Activity Report Lille - Nord Europe 2016

Edition: 2017-06-08

List of Inria's Research Teams

Project-Team BONSAI

Bioinformatics and Sequence Analysis

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
CNRS

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Computational Biology

Table of contents

Project-Team BONSAI

Creation of the Project-Team: 2011 January 01

Keywords:

Computer Science and Digital Science:
6.2.7. - High performance computing
7.2. - Discrete mathematics, combinatorics
7.9. - Graph theory

Other Research Topics and Application Domains:
1.1.6. - Genomics
1.1.7. - Immunology
1.1.8. - Evolutionnary biology
1.1.9. - Bioinformatics
1.1.13. - Plant Biology
1.1.14. - Microbiology
1.2. - Ecology
1.2.1. - Biodiversity
2.2.3. - Cancer

1. Members

Research Scientists
Hélène Touzet [Team leader, CNRS, Senior Researcher, HDR]
Samuel Blanquart [Inria, Researcher]
Rayan Chikhi [CNRS, Researcher]
Mathieu Giraud [CNRS, Researcher, HDR]

Faculty Members
Stéphane Janot [Univ. Lille I, Associate Professor]
Valérie Leclère [Univ. Lille I, Associate Professor, until Aug 2016, HDR]
Laurent Noé [Univ. Lille I, Associate Professor]
Maude Pupin [Univ. Lille I, Associate Professor, HDR]
Mikaël Salson [Univ. Lille I, Associate Professor]
Jean-Stéphane Varré [Univ. Lille I, Professor, HDR]

Technical Staff
Areski Flissi [CNRS]
Isabelle Guigon [CNRS, until Mar 2016]
Ryan Herbert [Inria]
Aurélien Béliard [CHU Lille, from Dec 2016]
Juraj Michalik [CNRS, until Mar 2016]

PhD Students
Yoann Dufresne [Univ. Lille I]
Pierre Marijon [Inria, from oct 2016]
Pierre Pericard [Univ. Lille I]
Tatiana Rocher [Univ. Lille I]
Chadi Saad [Univ. Lille II]

Léa Siegwald [CIFRE Gènes Diffusion]
Christophe Vroland [CNRS, until May 2016]

Post-Doctoral Fellow

Benjamin Momège [Inria]

Administrative Assistant

Amélie Supervielle [Inria]

2. Overall Objectives

2.1. Presentation

BONSAI is an interdisciplinary project whose scientific core is the design of efficient algorithms for the analysis of biological macromolecules.

From a historical perspective, research in bioinformatics started with string algorithms designed for the comparison of sequences. Bioinformatics became then more diversified and by analogy to the living cell itself, it is now composed of a variety of dynamically interacting components forming a large network of knowledge: Systems biology, proteomics, text mining, phylogeny, structural biology, etc. Sequence analysis still remains a central node in this interconnected network, and it is the heart of the BONSAI team.

It is a common knowledge nowadays that the amount of sequence data available in public databanks grows at an exponential pace. Conventional DNA sequencing technologies developed in the 70's already permitted the completion of hundreds of genome projects that range from bacteria to complex vertebrates. This phenomenon is dramatically amplified by the recent advent of Next Generation Sequencing (NGS), that gives rise to many new challenging problems in computational biology due to the size and the nature of raw data produced. The completion of sequencing projects in the past few years also teaches us that the functioning of the genome is more complex than expected. Originally, genome annotation was mostly driven by protein-coding gene prediction. It is now widely recognized that non-coding DNA plays a major role in many regulatory processes. At a higher level, genome organization is also a source of complexity and have a high impact on the course of evolution.

All these biological phenomena together with big volumes of new sequence data provide a number of new challenges to bioinformatics, both on modeling the underlying biological mechanisms and on efficiently treating the data. This is what we want to achieve in BONSAI. For that, we have in mind to develop well-founded models and efficient algorithms. Biological macromolecules are naturally modeled by various types of discrete structures: String, trees, and graphs. String algorithms is an established research subject of the team. We have been working on spaced seed techniques for several years. Members of the team also have a strong expertise in text indexing and compressed index data structures, such as BWT. Such methods are widely-used for the analysis of biological sequences because they allow a data set to be stored and queried efficiently. Ordered trees and graphs naturally arise when dealing with structures of molecules, such as RNAs or non-ribosomal peptides. The underlying questions are: How to compare molecules at structural level, how to search for structural patterns ? String, trees and graphs are also useful to study genomic rearrangements: Neighborhoods of genes can be modeled by oriented graphs, genomes as permutations, strings or trees.

A last point worth mentioning concerns the dissemination of our work to the biology and health scientific community. Since our research is driven by biological questions, most of our projects are carried out in collaboration with biologists. A special attention is given to the development of robust software, its validation on biological data and its availability from the software platform of the team: http://bioinfo.lille.inria.fr/.

3. Research Program

3.1. Sequence processing for Next Generation Sequencing

As said in the introduction of this document, biological sequence analysis is a foundation subject for the team. In the last years, sequencing techniques have experienced remarkable advances with Next Generation Sequencing (NGS), that allow for fast and low-cost acquisition of huge amounts of sequence data, and outperforms conventional sequencing methods. These technologies can apply to genomics, with DNA sequencing, as well as to transcriptomics, with RNA sequencing. They promise to address a broad range of applications including: Comparative genomics, individual genomics, high-throughput SNP detection, identifying small RNAs, identifying mutant genes in disease pathways, profiling transcriptomes for organisms where little information is available, researching lowly expressed genes, studying the biodiversity in metagenomics. From a computational point of view, NGS gives rise to new problems and gives new insight on old problems by revisiting them: Accurate and efficient remapping, pre-assembling, fast and accurate search of non exact but quality labeled reads, functional annotation of reads, ...

3.2. Noncoding RNA

Our expertise in sequence analysis also applies to noncoding RNA. Noncoding RNA plays a key role in many cellular processes. First examples were given by microRNAs (miRNAs) that were initially found to regulate development in *C. elegans*, or small nucleolar RNAs (snoRNAs) that guide chemical modifications of other RNAs in mammals. Hundreds of miRNAs are estimated to be present in the human genome, and computational analysis suggests that more than 20% of human genes are regulated by miRNAs. To go further in this direction, the 2007 ENCODE Pilot Project provides convincing evidence that the Human genome is pervasively transcribed, and that a large part of this transcriptional output does not appear to encode proteins. All those observations open a universe of “RNA dark matter” that must be explored. From a combinatorial point of view, noncoding RNAs are complex objects. They are single stranded nucleic acid sequences that can fold forming long-range base pairings. This implies that RNA structures are usually modeled by complex combinatorial objects, such as ordered labeled trees, graphs or arc-annotated sequences.

3.3. Genome structures

Our third application domain is concerned with the structural organization of genomes. Genome rearrangements are able to change genome architecture by modifying the order of genes or genomic fragments. The first studies were based on linkage maps and fifteen year old mathematical models. But the usage of computational tools was still limited due to the lack of data. The increasing availability of complete and partial genomes now offers an unprecedented opportunity to analyze genome rearrangements in a systematic way and gives rise to a wide spectrum of problems: Taking into account several kinds of evolutionary events, looking for evolutionary paths conserving common structure of genomes, dealing with duplicated content, being able to analyze large sets of genomes even at the intraspecific level, computing ancestral genomes and paths transforming these genomes into several descendant genomes.

3.4. Nonribosomal peptides

Lastly, the team has been developing for several years a tight collaboration with ProBioGEM team in Institut Charles Viollette on nonribosomal peptides, and has became a leader on that topic. Nonribosomal peptide synthesis produces small peptides not going through the central dogma. As the name suggests, this synthesis uses neither messenger RNA nor ribosome but huge enzymatic complexes called nonribosomal peptide synthetases (NRPSs). This alternative pathway is found typically in bacteria and fungi. It has been described for the first time in the 70’s. For the last decade, the interest in nonribosomal peptides and their synthetases has considerably increased, as witnessed by the growing number of publications in this field. These peptides are or can be used in many biotechnological and pharmaceutical applications (e.g. anti-tumors, antibiotics, immuno-modulators).

4. Application Domains

4.1. Life Sciences and health

Our research plays a pivotal role in all fields of life sciences and health where genomic data are involved. This includes more specifically the following topics: plant genomics (genome structure, evolution, microRNAs), cancer (leukemia, mosaic tumors), drug design (NRPSs), environment (metagenomics and metatranscriptomics), virology (evolution, RNA structures) ...

5. Highlights of the Year

5.1. Highlights of the Year

The software SortMeRNA, developed by the team, has reached the number of 100 labs worldwide that have been using it to analyze their sequencing data. SortMeRNA is able to deal with large metagenomics projects with multiple applications in health (gut microbiome,...), environment (sea, lakes, soil,...), biotechnologies (bio-films,...). The first version was released at the end of 2012, and it is still under active maintenance.

6. New Software and Platforms

6.1. BCALM 2

KEYWORDS: Bioinformatics - NGS - Genomics - Metagenomics - De Bruijn graphs
SCIENTIFIC DESCRIPTION

BCALM 2 is a bioinformatics tool for constructing the compacted de Bruijn graph from sequencing data. It is a parallel algorithm that distributes the input based on a minimizer hashing technique, allowing for good balance of memory usage throughout its execution. It is able to compact very large datasets, such as spruce or pine genome raw reads in less than 2 days and 40 GB of memory on a single machine.
FUNCTIONAL DESCRIPTION

BCALM 2 is an open-source tool for dealing with DNA sequencing data. It constructs a compacted representation of the de Bruijn graph. Such a graph is useful for many types of analyses, i.e. de novo assembly, de novo variant detection, transcriptomics, etc. The software is written in C++ and makes extensive use of the GATB library.

- Participants: Rayan Chikhi, Antoine Limasset and Paul Medvedev
- Contact: Rayan Chikhi
- URL: https://github.com/GATB/bcalm

6.2. NORINE

Nonribosomal peptides resource
KEYWORDS: Bioinformatics - Biotechnology - Biology - Genomics - Graph algorithmics - Chemistry - Knowledge database - Drug development - Computational biology
SCIENTIFIC DESCRIPTION

Since its creation in 2006, Norine remains the unique knowledgebase dedicated to non-ribosomal peptides (NRPs). These secondary metabolites, produced by bacteria and fungi, harbor diverse interesting biological activities (such as antibiotic, antitumor, siderophore or surfactant) directly related to the diversity of their structures. The Norine team goal is to collect the NRPs and provide tools to analyze them efficiently. We have developed a user-friendly interface and dedicated tools to provide a complete bioinformatics platform. The knowledgebase gathers abundant and valuable annotations on more than 1100 NRPs. To increase the quantity of described NRPs and improve the quality of associated annotations, we are now opening Norine to crowdsourcing. We believe that contributors from the scientific community are the best experts to annotate the NRPs they work on. We have developed MyNorine to facilitate the submission of new NRPs or modifications of stored ones. Norine is freely accessible from the following URL: http://bioinfo.lifl.fr/NRP.
FUNCTIONAL DESCRIPTION

Norine is a public computational resource with a web interface and REST access to a knowledge-base of nonribosomal peptides. It also contains dedicated tools : 2D graph viewer and editor, comparison of NRPs, MyNorine, a tool allowing anybody to easly submit new nonribosomal peptides, Smiles2monomers (s2m), a tool that deciphers the monomeric structure of polymers from their chemical structure.

- Participants: Maude Pupin, Areski Flissi, Valerie Leclere, Laurent Noe, Yoann Dufresne, Juraj Michalik and Stéphane Janot
- Partners: CNRS - Institut Charles Viollette - Université Lille 1 v
- Contact: Maude Pupin
- URL: http://bioinfo.lille.inria.fr/NRP

6.3. Olo

KEYWORDS: Bioinformatics - Indexation - Sequence alignment - Biological sequences - Approximate string matching
SCIENTIFIC DESCRIPTION

Approximate string matching of short sequences in a text often starts by a filtering step. That step relies on seed searching, which are shorter than the pattern. Usually in those seeds the number of errors is constrained, to allow more efficient computations. We designed the 01*0 seeds which offer a good trade-off between the number of false positives and filtering time.
FUNCTIONAL DESCRIPTION

We applied the 01*0 seeds to the similarity search of miRNA targets in a reference genome (Bwolo software) and to the similarity search between a pre-miRNA and mature miRNAs (Piccolo software).

- Participants: Sébastien Bini, Mikaël Salson, Hélène Touzet and Christophe Vroland
- Partners: CNRS - Université Lille 1
- Contact: Helene Touzet
- URL: http://bioinfo.lifl.fr/olo/

6.4. Vidjil

High-Throughput Analysis of V(D)J Immune Repertoire
KEYWORDS: Bioinformatics - NGS - Indexation - Cancer - Drug development
SCIENTIFIC DESCRIPTION

Vidjil is made of three components: an algorithm, a visualisation browser and a server that allow an analysis of lymphocyte populations containing V(D)J recombinations.

Vidjil high-throughput algorithm extracts V(D)J junctions and gather them into clones. This analysis is based on a spaced seed heuristics and is fast and scalable, as, in the first phase, no alignment is performed with database germline sequences. Each sequence is put in a cluster depending on its V(D)J junction. Then a representative sequence of each cluster is computed in time linear in the size of the cluster. Finally, we perform a full alignment using dynamic programming of that representative sequence against the germline sequences.

Vidjil also contains a dynamic browser (with D3JS) for visualization and analysis of clones and their tracking along the time in a MRD setup or in an immunological study.

FUNCTIONAL DESCRIPTION

Vidjil is an open-source platform for the analysis of high-throughput sequencing data from lymphocytes. V(D)J recombinations in lymphocytes are essential for immunological diversity. They are also useful markers of pathologies, and in leukemia, are used to quantify the minimal residual disease during patient follow-up. High-throughput sequencing (NGS/HTS) now enables the deep sequencing of a lymphoid population with dedicated Rep-Seq methods and software.

- Participants: Mathieu Giraud, Mikaël Salson, Marc Duez, Ryan Herbert, Tatiana Rocher and Florian Thonier
- Partners: CHRU Lille - CNRS - Inria - Université de Lille
- Contact: Mathieu Giraud
- URL: http://www.vidjil.org

7. New Results

7.1. Approximate pattern matching

The problem of measuring the similarity between two strings arises in many areas of sequence analysis. A common metric for it is the *Levenshtein distance*. This distance is defined as the smallest number of substitutions, insertions, and deletions of symbols required to transform one of the words into the other. We have investigated the basic problem of the size of the neighborhood of a given pattern P: count how many strings are within a bounded distance of a fixed reference string. There has been no efficient algorithm for calculating it so far. We have proposed a dynamic programming algorithm that scales linearly with the size of the pattern P. For that, we have introduced a new variant of the universal Levenshtein automaton, that is interesting by itself and that can have many other applications in text algorithms [31].

We have also addressed the related problem of approximate pattern matching: Given a text T and a pattern P, find all locations in T that differ by at most k errors (in the sense of the Levenshtein distance) from P. We have proposed a new kind of seeds (the 01^*0 seeds) that combines exact parts and parts with a fixed number of errors, and that are specifically well-suited for short DNA motifs with high error-rate. We have demonstrated the applicability of those seeds on two main case studies : pattern matching on a genomic scale with a Burrows-Wheeler transform, and multi-pattern matching with indexation of the set of patterns [30].

7.2. Parallel algorithm for de Bruijn graph compaction

Constructing a *de Bruijn graph* is an important step in the analysis of NGS data. This data structure is used in several applications, such as *de novo* assembly, variant detection, and transcriptome quantification. However, the representation of this graph often consumes prohibitive amounts of memory for large datasets. An operation, called compaction, enables to represent the graph more efficiently. However, so far, there was no algorithm for compacting the graph quickly and in low memory.

Along with colleagues at Inria Rennes and at Penn State University, we introduced a parallel algorithm and an implementation, BCALM 2, for constructing directly a compacted de Bruijn graph given a set of reads. Our results show that this algorithm enables to construct the graph for very large datasets, such as the spruce and pine genomes, in reasonable time and memory on a single machine. This represents a performance improvement of two orders of magnitude compared to previously available methods. BCALM 2 is open-source and was published at ISMB 2016 [20].

7.3. Range minimum query

The *range minimum query* problem consists in finding the minimum value inside any queried range of a preprocessed integer sequence. Several methods exist to compute the minimum in constant time, using almost the theoretical minimal amount of space. Those methods consist in splitting the problem in several subproblems and precomputing the solutions for them.

With Alice Héliou (AMIB Inria team, Saclay), Martine Léonard and Laurent Mouchard (LITIS, Rouen), we designed a new method, which is worse in terms of time complexity [24]. Our solution relies on a totally different concept as previous ones: We only store the values that are local minima. This approach is therefore simple and can, on specific inputs, require much less memory than the general theoretical minimal bound. Moreover the simplicity of the method can be easily adapted to allow updates in the original integer sequence.

7.4. Coding isoform structures

Our researches on gene isoform structures started in 2014 with the CG-Alcode Associated Team and in collaboration with Anne Bergeron from the LACIM (Montréal, Canada). We aimed at defining better definitions of isoform orthology at the coding level, which are based on the preservation of all the exon junctions in two orthologous isoforms. This estimation is achieved at the gene level, where sequence homology is detected for both exons and their flanking intronic splice sites [19]. The approach largely outperforms competing programs in terms of precision and recall. Using the successive releases of the ten years old CCDS database, we show that the discovery rate of orthologous isoforms between human and mouse is growing continuously and that it displays no sign of completion.

7.5. Nonribosomal peptides

We were invited to contribute in a volume of “Methods in Molecular Biology” by authoring a chapter focusing on NRPS biosynthesis. This chapter [32] was about the use of the Norine platform (developed by the team) and other bioinformatics tools for the analysis of nonribosomal peptide synthetases and their products. We invited our collaborator from Denmark, Tilmann Weber, to complete this chapter with the introduction of his tool, antiSMASH.

We annotated 48 genomes of *Burkholderia* species using our annotation protocol, that starts from a genome sequence and goes to the predicted nonribosomal peptides. We have predicted 161 gene clusters producing nonribosomal peptides, leading to the synthesis of not only already known peptides, but also new ones [22] with potential applications in biocontrol.

A new version of the Norine interface is now available. The form to query the annotations is now flexible and dynamic. The user can build his own query to search for annotations in several fields combined by boolean operators. Moreover, the database structure has been modified to allow, among others, a hierarchical representation of the NRPS taxonomy. Finally, the MyNorine tool has been enhanced and updated to take into account these changes and the description page of the peptides has been reorganized.

7.6. High-throughput V(D)J repertoire analysis

Researches on high-throughput V(D)J repertoire analysis started in the group in 2012. We have developed Vidjil, a web platform dedicated to the analysis of lymphocyte populations. Starting from DNA sequences, uploaded by the user, Vidjil identifies and quantifies lymphocyte populations and provides an interactive visualization [21].

In 2016, with our colleagues at Lille hospital, we published two articles in haematological journals to detail our method for the diagnosis [23] and for the follow-up [28] of the acute lymphoblastic leukemia using high-throughput sequencing. Our results also show what those new techniques, together with bioinformatics software, bring in a routine practice. Being a full platform with metadata storage, Vidjil is used on a regular basis by about 20 laboratories around the world. In France, the majority of diagnosis samples from acute lymphoblastic leukemia patients are now analyzed using Vidjil.

7.7. Assembly of the giraffe genome and the gorilla Y-chromosome

We collaborated with two labs from the Pennsylvania State Institute (Cavener Lab and Makova Lab) for practical analysis of DNA sequencing data. The first collaboration led to the publication of the giraffe genome in Nature Communication [18]. In this article our contribution was to provide the first draft-quality whole-genome sequences of the giraffe and the okapi. The second collaboration was about assembling the Y-chromosome of the gorilla using a novel sequencing strategy as well as novel computational tools. This work was published in Genome Research [29].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

The PhD thesis of Léa Siegwald is funded by a CIFRE contract with the biotechnology company Gènes Diffusion.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

- ANR ASTER: ASTER is a national project that aims at developing algorithms and software for analyzing third-generation sequencing data, and more specifically RNA sequencing. BONSAI is the principal investigator in this ANR. Other partners are Erable (LBBE in Lyon) and two sequencing and analysis platforms that have been very active in the MinION Access Program (Genoscope and Institut Pasteur de Lille).
- PIA France Génomique: National funding from "Investissements d'Avenir" (call *Infrastructures en Biologie-Santé*). France Génomique is a shared infrastructure, whose goal is to support sequencing, genotyping and associated computational analysis, and increases French capacities in genome and bioinformatics data analysis. It gathers 9 sequencing and 8 bioinformatics platforms. Within this consortium, we are responsible for the workpackage devoted to the computational analysis of sRNA-seq data, in coordination with the bioinformatics platform of Génopole Toulouse-Midi-Pyrénées.

9.1.2. ADT

- ADT Vidjil (2015–2017): The purpose of this ADT is to strengthen Vidjil development and to ensure a better diffusion of the software by easing its installation, administration and usability. This will enable the software to be well suited for a daily clinical use. The software is already used in test on our own web server (more than 5,000 samples processed by now). Vidjil is now used in a routine practice by three French hospitals and one German hospital. By the end of the ADT, we expect this number to increase and the software to be directly installed inside some hospitals.

9.2. European Initiatives

9.2.1. Collaborations in European Programs, Except FP7 & H2020

International ANR RNAlands (2014-2017): National funding from the French Agency Research (call *International call*). Our objective is the fast and efficient sampling of structures in RNA Folding Landscapes. The project gathers three partners: Amib from Inria Saclay, the Theoretical Biochemistry Group from Universität Wien and BONSAI.

Interreg Va (France-Wallonie-Vlaanderen) : Portfolio "SmartBioControl", including 5 constitutive projects and 25 partners working together towards sustainable agriculture.

9.3. International Initiatives

9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

9.3.1.1. CG-ALCODE

Title: Comparative Genomics for the analysis of gene structure evolution: ALternative CODing in Eukaryote genes through alternative splicing, transcription, and translation.

International Partner (Institution - Laboratory - Researcher):

Université du Québec À Montréal (Canada) - Laboratoire de combinatoire et d'informatique mathématique (LaCIM) - Anne Bergeron

From 2014 to 2017

The aim of this Associated Team is the development of comparative genomics models and methods for the analysis of eukaryote genes structure evolution. Our goal is to answer very important questions arising from recent discoveries on the major role played by alternative transcription, splicing, and translation, in the functional diversification of eukaryote genes. Two working meeting took place in 2016. S. Blanquart and J.-S. Varré met A. Bergeron and K. Swenson in Montpellier, from 13th to 15th of April. J.-S. Varré and K. Swenson met A. Bergeron in Montréal, from 8th to 19th of November.

9.3.2. Inria International Partners

9.3.2.1. Informal International Partners

- *Astrid Lindgrens Hospital, Stockholm University:* Collaboration with Anna Nilsson and Shanie Saghafian-Hedengren on RNA sequencing of stromal cells.
- *Childhood Leukaemia Investigation Prague (CLIP), Department of Pediatric Hematology/Oncology, 2nd Faculty of Medicine, Charles University, Prague, Czech Republic:* Collaboration with Michaela Kotrová and Eva Fronkova on leukemia diagnosis and follow-up.
- *CWI Amsterdam:* Collaboration with Alexander Schoenhuth and Jasmijn Baaijens on succinct data structures and algorithms for the assembly of viral quasispecies.
- *Department of Statitics, North Carolina State University:* Collaboration with Donald E. K. Martin on spaced seeds coverage.
- *Département des Sciences de la Vie, Faculté des Sciences de Liège:* Collaboration with Denis Beaurain on nonribosomal peptides.
- *Gembloux Agro-Bio Tech, Université de Liège:* Collaboration with Philippe Jacques on nonribosomal peptides.
- *Institut für Biophysik und physikalische Biochemie, University of Regensburg:* Collaboration with Rainer Merkl on ancestral sequence inference and synthesis.
- *Institute of Biosciences and Bioresources, Bari:* Collaboration with Nunzia Scotti on the assembly of plant mitochondrial genomes.
- *Makova lab, The Pennsylvania State University:* Collaboration with Kateryna Makova and Samarth Rangavittal on the assembly of the gorilla Y chromosome, and visualisation of assembly graphs.
- *Medvedev lab, The Pennsylvania State University:* Collaboration with Paul Medvedev on algorithms for constructing de Bruijn graphs.

- *Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark:* Collaboration with Tilmann Weber on nonribosomal peptides.
- *Proteome Informatics Group, Swiss Institute of Bioinformatics:* Collaboration with Frédérique Lisacek on nonribosomal peptides.
- *School of Social and Community Medicine, University of Bristol:* Collaboration with Marc Duez, John Moppett and Stephanie Wakeman on leukemia diagnosis follow-up.
- *Theoretical Biochemistry Group, Universität Wien:* Collaboration with Andrea Tanzer and Ronny Lorenz on RNA folding and RNA kinetics.

9.3.3. Participation in Other International Programs

- Participation in the EuroClonality-NGS consortium. This consortium aims at standardizing the study of immune repertoire, clonality and minimal residual disease in leukemia at the european level. We are part of the bioinformatics workgroup led by Nikos Darzentas (CEITEC, Brno, Czech Republic).

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Dr. Alexander Schoenhuth (group leader, *CWI Amsterdam*) and Jasmijn Baaijens (PhD student, *CWI Amsterdam*).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organization

10.1.1.1. General Chair, Scientific Chair

- RepSeq 2016, workshop on immune repertoire sequencing at ECCB 2016 (M. Giraud, M. Salson).

10.1.1.2. Member of the Organizing Committees

- SMPGD, Statistical Methods in Post Genomic Data (H. Touzet).

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- WABI 2016 (H. Touzet).
- RECOM-seq 2016 (H. Touzet).

10.1.2.2. Reviewer

- ECCB 2016 (J.-S. Varré, R. Chikhi).
- PSC 2016 (M. Salson).
- RECOMB-CG 2016 (J.-S. Varré).
- TCBB 2016 (J.-S. Varré).
- RECOMB 2016 (R. Chikhi).

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- Bioinformatics (M. Salson, J.-S. Varré, R. Chikhi, H. Touzet).
- PLoS Genetics (R. Chikhi).
- Genome Research (R. Chikhi).

- Nucleic Acids Research (R. Chikhi).
- Journal of Discrete Algorithms (M. Salson).
- Briefing in Bioinformatics (H. Touzet).

10.1.4. Invited Talks

- RegPep2016 (regulated peptides) in Rouen (France), Symposium 8 "In silico approaches to peptide identification and design" (M. Pupin).
- SMPGD keynote in Lille (France), (R. Chikhi).
- National meeting of GDR Informatique Mathématique in Paris (H. Touzet).
- MatBio 2016 in London (H. Touzet).
- Summer school in metagenomics in Paris (H. Touzet).

10.1.5. Scientific Expertise

- Reviewer for a Swiss National Science Foundation grant (R. Chikhi).
- Member of the scientific committee of the national program Environmics (H. Touzet).
- Reviewer for a CRSNG grant (H. Touzet).
- Reviewer for labex CIMI PhD grant (H. Touzet).
- Member of three HCERES committees: L2N (LINA and IRCCyN, Nantes), Loria (Nancy), I2M (Marseille) (H. Touzet).

10.1.6. Research Administration

- Member of the CUB for Inria Lille (S. Blanquart).
- Member of the Charles Viollette Institute Laboratory council (V. Leclère).
- Member of the Charles Viollette Institute scientific committee (V. Leclère).
- Member of the scientific operational committee of Xperium, Univ. Lille 1 (V. Leclère).
- Member of the Inria local committee for technology development (M. Pupin).
- Member of the executive council of the IFB, Institut Français de Bioinformatique, (M. Pupin).
- Member of the Inria local committee for the IT users (M. Salson).
- Member of the national scientific committee of INS2I–CNRS (H. Touzet).
- Member of the scientific committee of MBIA – INRA (H. Touzet).
- Head of the national CNRS network GDR Bioinformatique moléculaire (http://www.gdr-bim.cnrs.fr, H. Touzet).
- Vice-head of the Lille Bioinformatics platform, bilille (H. Touzet).
- Member of the CRIStAL Laboratory council (H. Touzet).
- Member of the CRIStAL scientific council, coordinator of the thematic group “Modeling for life sciences” (J.-S. Varré).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Teaching in computer science:

- Master: Y. Dufresne, *Algorithmics and complexity*, 36h, M1 Computer Science, Univ. Lille 1.
- License: Y. Dufresne, *Oriented object design*, 42h, L3 Computer Science, Univ. Lille 1.
- License: S. Janot, *Introduction to programming (C)*, 50h, L3 Polytech’Lille, Univ. Lille 1.
- License: S. Janot, *Databases*, 30h, L3 Polytech’Lille, Univ. Lille 1.

- Master: S. Janot, *Databases*, 12h, M1 Polytech'Lille, Univ. Lille 1.
- Master: S. Janot, *Logic and Semantic Web*, 80h, M1 Polytech'Lille, Univ. Lille 1.
- License: L. Noé, *Networks*, 42h, L3 Computer science, Univ. Lille 1.
- License: L. Noé, *Programming (Python)*, 54h, L3 Computer science' S3H, Univ. Lille 1.
- License: L. Noé, *Coding and information theory*, 36h, L2 Computer science, Univ. Lille 1.
- License: L. Noé, *Functional Programming*, 30h, L2 Computer science, Univ. Lille 1.
- License: P. Pericard, *Data structures*, 18h, L3 Polytech'Lille, Univ. Lille 1.
- License: P. Pericard, *Introduction to programming (C)*, 34h, L3 Polytech'Lille, Univ. Lille 1.
- License: P. Pericard, *Programming (C)*, 22h, L3 Polytech'Lille, Univ. Lille 1.
- License: P. Pericard, *Databases*, 22h, M1 Polytech'Lille, Univ. Lille 1.
- License: M. Pupin, *Introduction to programming (Python)*, 78h, L1 Computer science, Univ. Lille 1.
- License: M. Pupin, *Professional project*, 18h, L3 Computer science, Univ. Lille 1.
- Master: M. Pupin, *Introduction to programming (JAVA)*, 24h, M1 Mathématiques et finance, Univ. Lille 1.
- License: T. Rocher, *Algorithmics and programming*, 32h, L3 Polytech'Lille, Univ. Lille 1.
- License: T. Rocher, *Algorithmics and programming (remedial course)*, 8h, L3 Polytech'Lille, Univ. Lille 1.
- License: T. Rocher, *Databases*, 26h, L3 Polytech'Lille, Univ. Lille 1.
- License: C. Saad, *Algorithmics and programming*, 28h, L3 Polytech'Lille, Univ. Lille 1.
- License: C. Saad, *Databases*, 36h, L3 Polytech'Lille, Univ. Lille 1.
- Master: M. Salson, *Skeptical thinking*, 18h, M2 Journalist and Scientist, ESJ, Univ. Lille 1.
- License: M. Salson, *Coding and information theory*, 63h, L2 Computer science, Univ. Lille 1.
- License: J.-S. Varré, *Web programming*, 36h, L2 Computer Science, Univ. Lille 1.
- License: J.-S. Varré, Y. Dufresne, *Object oriented programming*, 36h, L2 Computer Science, Univ. Lille 1.
- License: J.-S. Varré, *Algorithms and data structures*, 50h, L2 Computer science, Univ. Lille 1.
- License: J.-S. Varré, *System*, 36h, L3 Computer science, Univ. Lille 1.
- Master: J.-S. Varré, Y. Dufresne, *Software project*, 24h, M1 Computer science, Univ. Lille 1.

Teaching in bioinformatics:

- License: S. Blanquart, R. Chikhi, M. Giraud, *Bioinformatics*, 40h, L3 Computer Science, Univ. Lille 1.
- Master: S. Blanquart, *Algorithms and applications in bioinformatics*, 24h, M1 Computer Science, Univ. Lille 1.
- Master: S. Blanquart, *Methods in phylogenetics*, 4h, M2 Biodiversité Evolution Ecologie, Univ. Lille 1.
- License: V. Leclère, *Biotechnology*, 24h, L3 Biology, Univ. Lille 1.
- Master: L. Noé, *Bioinformatics*, 40h, M1 Biotechnologies, Univ. Lille 1.
- Master: M. Pupin *Bioinformatics*, 40h, M1 Biology and Biotechnologies, Univ. Lille 1.
- Master: M. Salson, *Algorithms for life sciences*, 18h, M2 Complex models, algorithms and data, Univ. Lille 1.

Teaching in biology:

- Master, V. Leclère, *Mycology, secondary metabolites, food microbiology*, 37 h, M1 Biology, Univ. Lille 1.

10.2.2. Teaching administration

- Head of the computer science modules in the 1st year of licence, univ. Lille 1 (M. Pupin).
- Head of the licence semester “Computer Science – S3 Harmonisation (S3H)”, univ. Lille 1 (L. Noé).
- Member of UFR IEEA council (M. Pupin, J.-S. Varré).
- Head of the 3rd year of licence of computer science, univ. Lille 1(J.-S. Varré).
- Head of the GIS department (Software Engineering and Statistics) of Polytech’Lille (S. Janot).
- Member of UFR Biologie council (V. Leclère).
- Head of the master “Innovations en biotechnologies végétales, enzymatiques et microbiennes”, univ. Lille 1 (V. Leclère).

10.2.3. Supervision

- PhD : C. Vroland, Algorithmique pour la recherche de motifs approchée et application à la recherche de cibles de microARN, univ. Lille 1, 2016/05/18, H. Touzet, V. Castric, M. Salson.
- PhD : Y. Dufresne, Algorithmes pour l’annotation automatique de peptides non-ribosomiques, univ. Lille 1, 2016/12/01, M. Pupin, L. Noé.
- PhD in progress: P. Pericard, Methods for taxonomic assignation in metagenomics, 2013/11/01, H. Touzet, S. Blanquart.
- PhD in progress: T. Rocher, Indexing VDJ recombinations in lymphocytes for leukemia follow-up, 2014/11/01, M. Giraud, M. Salson.
- PhD in progress: C. Saad, Caractérisation des erreurs de séquençage non aléatoires, application aux mosaïques et tumeurs hétérogènes, 2014/10/01, M.-P. Buisine, H. Touzet, J. Leclerc, L. Noé, M. Figeac.
- PhD in progress: L. Siegwald, Bionformatic analysis of Ion Torrent metagenomic data, 2014/01/03, H. Touzet, Y. Lemoine (Institut Pasteur de Lille).
- PhD in progress: P. Marijon, Graph assembly analysis for third generation sequencing data, 2016/10/01, J.-S. Varré, R. Chikhi (Institut Pasteur de Lille).

10.2.4. Juries

- Member of the HDR committee of Laurent Mouchard (Univ. Rouen, J.-S. Varré).
- Member of the PhD committee of Qassin Esmael (Univ. Lille 1, M. Pupin, V. Leclère).
- Member of the PhD committee of Souhir Sabri (Univ Montpellier, V. Leclère).
- Member of the PhD committee of Wahiba Chaara (Univ. Paris 6, M. Giraud).
- Member of the PhD committee of Leandro Ishi (Univ. Lyon, R. Chikhi).
- Member of the PhD committee of Jerome Audoux (Univ. Montpellier, R. Chikhi, M. Salson).
- Member of the PhD jury of Clara Benoit (Univ. Lyon, R. Chikhi).
- Member of the HDR jury of Morgane Thomas-Chollier (IBENS, Ecole Normale Supérieure Paris, H. Touzet).
- Member of the HDR jury of Annie Chateau (LIRMM, Université de Montpellier, H. Touzet).
- Member of the HDR jury of Pierre Peterlongo (Inria Rennes, H. Touzet).
- Member of the HDR jury of Séverine Bérard (ISEM, Université de Montpellier, H. Touzet).
- Member of the PhD jury of Thomas Hume (LaBRI, Université Bordeaux 1, H. Touzet).
- Member of the PhD jury of Aymeric Antoine-Lorquin (IRISA, Université de Rennes 1, H. Touzet).
- Member of the hiring committee MdC of Univ. Nancy (M. Pupin).
- Member of the hiring committee MdC of Univ Lille 1 (V. Leclère).
- Member of the hiring committee professor of Univ. Rouen (H. Touzet).

- Member of the hiring committee Research Engineer of Univ. Paris-Diderot (M. Pupin).
- Member of the hiring committee of research engineer of Univ Lille 1 (V. Leclère).

10.3. Popularization

The team has always been very active in popularizing computational biology and computer science in general.

- M. Salson participated in an exchange with scientific journalists organized by the French association of scientific journalists (AJSPI). A journalist was hosted during one week in the team and M. Salson spent one week in the newsroom of *La Recherche*, a French science magazine.
- Within a project on skeptical thinking with a popularization association “Les Petits Débrouillards”, M. Salson is part of the monitoring committee and gave lectures to social workers.
- The team participates to dissemination actions for high school students and high school teachers on a regular basis: multiple presentations on bioinformatics and research in bioinformatics with our dedicated “genome puzzles”, booth about computer science unplugged for high school girls, booth at Xperium about development of biopesticides (including a demo on the use of Norine), plenary presentations at the “Day for Programming and Algorithmic Teaching”, presentations at “Salon de l'étudiant”, visit of high school students in the team (M. Giraud, M. Pupin, M. Salson, J.-S. Varré, R. Chikhi, V. Leclère)

11. Bibliography

Major publications by the team in recent years

[1] A. ABDO, S. CABOCHE, V. LECLÈRE, P. JACQUES, M. PUPIN.*A new fingerprint to predict nonribosomal peptides activity*, in "Journal of Computer-Aided Molecular Design", October 2012, vol. 26, n^o 10, p. 1187-94 [*DOI :* 10.1007/S10822-012-9608-4], http://hal.inria.fr/hal-00750002.

[2] A. ABDO, V. LECLÈRE, P. JACQUES, N. SALIM, M. PUPIN.*Prediction of new bioactive molecules using a bayesian belief network*, in "Journal of Chemical Information and Modeling", January 2014, vol. 54, n^o 1, p. 30-36 [*DOI :* 10.1021/CI4004909], https://hal.archives-ouvertes.fr/hal-01090611.

[3] R. CHIKHI, A. LIMASSET, P. MEDVEDEV.*Compacting de Bruijn graphs from sequencing data quickly and in low memory*, in "Bioinformatics", November 2016, vol. 32, n^o 12, p. i201 - i208 [*DOI :* 10.1093/BIOINFORMATICS/BTW279], https://hal.archives-ouvertes.fr/hal-01395704.

[4] Y. DUFRESNE, L. NOÉ, V. LECLÈRE, M. PUPIN.*Smiles2Monomers: a link between chemical and biological structures for polymers*, in "Journal of Cheminformatics", December 2015 [*DOI :* 10.1186/S13321-015-0111-5], https://hal.inria.fr/hal-01250619.

[5] Y. FERRET, A. CAILLAULT, S. SEBDA, M. DUEZ, N. GRARDEL, N. DUPLOYEZ, C. VILLENET, M. FIGEAC, C. PREUDHOMME, M. SALSON, M. GIRAUD.*Multi-loci diagnosis of acute lymphoblastic leukaemia with high-throughput sequencing and bioinformatics analysis*, in "British Journal of Haematology", 2016, bjh.13981 [*DOI :* 10.1111/BJH.13981], https://hal.archives-ouvertes.fr/hal-01279160.

[6] A. FLISSI, Y. DUFRESNE, J. MICHALIK, L. TONON, S. JANOT, L. NOÉ, P. JACQUES, V. LECLÈRE, M. PUPIN.*Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing*, in "Nucleic Acids Research", 2015 [*DOI :* 10.1093/NAR/GKV1143], https://hal.archives-ouvertes.fr/hal-01235996.

[7] M. FRITH, L. NOÉ.*Improved search heuristics find 20 000 new alignments between human and mouse genomes*, in "Nucleic Acids Research", February 2014, vol. 42, n^o 7, e59 [*DOI :* 10.1093/NAR/GKU104], https://hal.inria.fr/hal-00958207.

[8] R. GIEGERICH, H. TOUZET.*Modeling dynamic programming problems over sequences and trees with inverse coupled rewrite systems*, in "Algorithms", 2014, vol. 7, p. 62 - 144 [*DOI :* 10.3390/A7010062], https://hal.archives-ouvertes.fr/hal-01084318.

[9] M. GIRAUD, M. SALSON, M. DUEZ, C. VILLENET, S. QUIEF, A. CAILLAULT, N. GRARDEL, C. ROUMIER, C. PREUDHOMME, M. FIGEAC.*Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing*, in "BMC Genomics", 2014, vol. 15, n^o 1, 409 [*DOI :* 10.1186/1471-2164-15-409], https://hal.archives-ouvertes.fr/hal-01009173.

[10] E. KOPYLOVA, L. NOÉ, H. TOUZET.*SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data*, in "Bioinformatics", October 2012, p. 1-10 [*DOI :* 10.1093/BIOINFORMATICS/BTS611], http://hal.inria.fr/hal-00748990.

[11] M. LÉONARD, L. MOUCHARD, M. SALSON.*On the number of elements to reorder when updating a suffix array*, in "Journal of Discrete Algorithms", February 2012, vol. 11, p. 87-99 [*DOI :* 10.1016/J.JDA.2011.01.002], http://hal.inria.fr/inria-00636066.

[12] D. E. K. MARTIN, L. NOÉ.*Faster exact distributions of pattern statistics through sequential elimination of states*, in "Annals of the Institute of Statistical Mathematics", September 2015 [*DOI :* 10.1007/S10463-015-0540-Y], https://hal.inria.fr/hal-01237045.

[13] L. NOÉ, D. E. K. MARTIN.*A coverage criterion for spaced seeds and its applications to support vector machine string kernels and k-mer distances*, in "Journal of Computational Biology", November 2014, vol. 21, n^o 12, 28 [*DOI :* 10.1089/CMB.2014.0173], https://hal.inria.fr/hal-01083204.

[14] A. PERRIN, J.-S. VARRÉ, S. BLANQUART, A. OUANGRAOUA.*ProCARs: progressive reconstruction of ancestral gene orders*, in "BMC Genomics", 2015, vol. 16, n^o Suppl 5, S6 [*DOI :* 10.1186/1471-2164-16-S5-S6], https://hal.inria.fr/hal-01217311.

[15] M. PUPIN, Q. ESMAEEL, A. FLISSI, Y. DUFRESNE, P. JACQUES, V. LECLÈRE.*Norine: a powerful resource for novel nonribosomal peptide discovery*, in "Synthetic and Systems Biotechnology", December 2015 [*DOI :* 10.1016/J.SYNBIO.2015.11.001], https://hal.inria.fr/hal-01250614.

[16] A. SAFFARIAN, M. GIRAUD, A. DE MONTE, H. TOUZET.*RNA locally optimal secondary structures*, in "Journal of Computational Biology", 2012, vol. 19, n^o 10, p. 1120-1133 [*DOI :* 10.1089/CMB.2010.0178], http://hal.inria.fr/hal-00756249.

[17] A. SAFFARIAN, M. GIRAUD, H. TOUZET.*Modeling alternate RNA structures in genomic sequences*, in "Journal of Computational Biology", February 2015, vol. 22, n^o 3, p. 190-204, https://hal.archives-ouvertes.fr/hal-01228130.

Publications of the year

Articles in International Peer-Reviewed Journal

[18] M. AGABA, E. ISHENGOMA, W. C. MILLER, B. C. MCGRATH, C. N. HUDSON, O. C. BEDOYA REINA, A. RATAN, R. BURHANS, R. CHIKHI, P. MEDVEDEV, C. A. PRAUL, L. CAVENER, B. WOOD, H. ROBERTSON, L. PENFOLD, D. R. CAVENER.*Giraffe genome sequence reveals clues to its unique morphology and physiology*, in "Nature Communications", May 2016, vol. 7 [*DOI :* 10.1038/NCOMMS11519], https://hal.archives-ouvertes.fr/hal-01395703.

[19] S. BLANQUART, J.-S. VARRÉ, P. GUERTIN, A. PERRIN, A. BERGERON, K. M. SWENSON.*Assisted transcriptome reconstruction and splicing orthology*, in "BMC Genomics", 2016, vol. 17, n° 786 [*DOI :* 10.1186/S12864-016-3103-6], https://hal.inria.fr/hal-01396410.

[20] R. CHIKHI, A. LIMASSET, P. MEDVEDEV.*Compacting de Bruijn graphs from sequencing data quickly and in low memory*, in "Bioinformatics", November 2016, vol. 32, n° 12, p. i201 - i208 [*DOI :* 10.1093/BIOINFORMATICS/BTW279], https://hal.archives-ouvertes.fr/hal-01395704.

[21] M. DUEZ, M. GIRAUD, R. HERBERT, T. ROCHER, M. SALSON, F. THONIER.*Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing*, in "PLoS ONE", November 2016, vol. 11, n° 11 [*DOI :* 10.1371/JOURNAL.PONE.0166126], https://hal.archives-ouvertes.fr/hal-01397079.

[22] Q. ESMAEEL, M. PUPIN, N. P. KIEU, G. CHATAIGNÉ, M. BÉCHET, J. DERAVEL, F. KRIER, M. HÖFTE, P. JACQUES, V. LECLÈRE.*Burkholderia genome mining for nonribosomal peptide synthetases reveals a great potential for novel siderophores and lipopeptides synthesis*, in "MicrobiologyOpen", May 2016, vol. 5, n° 3, p. 512 - 526 [*DOI :* 10.1002/MBO3.347], https://hal.archives-ouvertes.fr/hal-01398944.

[23] Y. FERRET, A. CAILLAULT, S. SEBDA, M. DUEZ, N. GRARDEL, N. DUPLOYEZ, C. VILLENET, M. FIGEAC, C. PREUDHOMME, M. SALSON, M. GIRAUD.*Multi-loci diagnosis of acute lymphoblastic leukaemia with high-throughput sequencing and bioinformatics analysis*, in "British Journal of Haematology", 2016, bjh.13981 [*DOI :* 10.1111/BJH.13981], https://hal.archives-ouvertes.fr/hal-01279160.

[24] A. HELIOU, M. LÉONARD, L. MOUCHARD, M. SALSON.*Efficient dynamic range minimum query*, in "Theoretical Computer Science", 2017 [*DOI :* 10.1016/J.TCS.2016.07.002], https://hal.archives-ouvertes.fr/hal-01255499.

[25] T. MARSCHALL, M. MARZ, T. ABEEL, L. DIJKSTRA, B. E. DUTILH, A. GHAFFAARI, P. KERSEY, W. P. KLOOSTERMAN, V. MAKINEN, A. M. NOVAK, B. PATEN, D. PORUBSKY, E. RIVALS, C. ALKAN, J. A. BAAIJENS, P. I. W. D. BAKKER, V. BOEVA, R. J. P. BONNAL, F. CHIAROMONTE, R. CHIKHI, F. D. CICCARELLI, R. CIJVAT, E. DATEMA, C. M. V. DUIJN, E. E. EICHLER, C. ERNST, E. ESKIN, E. GARRISON, M. EL-KEBIR, G. W. KLAU, J. O. KORBEL, E.-W. LAMEIJER, B. LANGMEAD, M. MARTIN, P. MEDVEDEV, J. C. MU, P. NEERINCX, K. OUWENS, P. PETERLONGO, N. PISANTI, S. RAHMANN, B. RAPHAEL, K. REINERT, D. D. RIDDER, J. D. RIDDER, M. SCHLESNER, O. SCHULZ-TRIEGLAFF, A. D. SANDERS, S. SHEIKHIZADEH, C. SHNEIDER, S. SMIT, D. VALENZUELA, J. WANG, L. WESSELS, Y. ZHANG, V. GURYEV, F. VANDIN, K. YE, A. SCHÖNHUTH.*Computational pan-genomics: status, promises and challenges*, in "Briefings in Bioinformatics", October 2016 [*DOI :* 10.1093/BIB/BBW089], https://hal.inria.fr/hal-01390478.

[26] D. E. K. MARTIN, L. NOÉ.*Faster exact distributions of pattern statistics through sequential elimination of states*, in "Annals of the Institute of Statistical Mathematics", February 2017, vol. 69, n° 1, p. 231–248 [*DOI :* 10.1007/S10463-015-0540-Y], https://hal.inria.fr/hal-01237045.

[27] K. SAHLIN, R. CHIKHI, L. ARVESTAD.*Assembly scaffolding with PE-contaminated mate-pair libraries*, in "Bioinformatics", March 2016, vol. 32, n° 13, p. 1925 - 1932 [*DOI :* 10.1093/BIOINFORMATICS/BTW064], https://hal.archives-ouvertes.fr/hal-01396904.

[28] M. SALSON, M. GIRAUD, A. CAILLAULT, N. GRARDEL, N. DUPLOYEZ, Y. FERRET, M. DUEZ, R. HERBERT, T. ROCHER, S. SEBDA, S. QUIEF, C. VILLENET, M. FIGEAC, C. PREUDHOMME.*High-throughput sequencing in acute lymphoblastic leukemia: Follow-up of minimal residual disease and emergence of new clones*, in "Leukemia Research", 2017, vol. 53, p. 1-7 [*DOI :* 10.1016/J.LEUKRES.2016.11.009], https://hal.archives-ouvertes.fr/hal-01404817.

[29] M. TOMASZKIEWICZ, S. RANGAVITTAL, M. CECHOVA, R. C. SANCHEZ, H. W. FESCEMYER, R. HARRIS, D. YE, P. C. O'BRIEN, R. CHIKHI, O. A. RYDER, M. A. FERGUSON-SMITH, P. MEDVEDEV, K. D. MAKOVA.*A time- and cost-effective strategy to sequence mammalian Y Chromosomes: an application to the de novo assembly of gorilla Y*, in "Genome Research", March 2016, vol. 26, n° 4, p. 530 - 540 [*DOI :* 10.1101/GR.199448.115], https://hal.archives-ouvertes.fr/hal-01395702.

[30] C. VROLAND, M. SALSON, S. BINI, H. TOUZET.*Approximate search of short patterns with high error rates using the 010 lossless seeds*, in "Journal of Discrete Algorithms", 2016, vol. 37, p. 3-16 [*DOI :* 10.1016/J.JDA.2016.03.002], https://hal.archives-ouvertes.fr/hal-01360485.

International Conferences with Proceedings

[31] H. TOUZET.*On the Levenshtein Automaton and the Size of the Neighborhood of a Word*, in "Language and Automata Theory and Applications", Prague, Czech Republic, A.-H. DEDIU, J. JANOUŠEK, C. MARTÍN-VIDE, B. TRUTHE (editors), Lecture Notes in Computer Sciences, Springer, 2016, vol. 9618, p. 207-218 [*DOI :* 10.1007/978-3-319-30000-9_16], https://hal.archives-ouvertes.fr/hal-01360482.

Scientific Books (or Scientific Book chapters)

[32] V. LECLÈRE, T. WEBER, P. JACQUES, M. PUPIN.*Bioinformatics Tools for the Discovery of New Nonribosomal Peptides*, in "Nonribosomal Peptide and Polyketide Biosynthesis", B. S. EVANS (editor), Methods in Molecular Biology, Humana Press, February 2016, vol. 1401, p. 209-232 [*DOI :* 10.1007/978-1-4939-3375-4_14], https://hal.archives-ouvertes.fr/hal-01398960.

Other Publications

[33] M. SALSON, A. CAILLAULT, M. DUEZ, Y. FERRET, A. FIEVET, M. KOTROVA, F. THONIER, P. VILLARESE, S. WAKEMAN, G. WRIGHT, M. GIRAUD.*A dataset of sequences with manually curated V(D)J designations*, 2016, Workshop Immune Repertoire Sequencing : Bioinformatics and Applications in Hematology and Immunology (RepSeq 2016), https://hal.archives-ouvertes.fr/hal-01331556.

Team DEFROST

Deformable Robotic Software

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Lille - Nord Europe

THEME
Robotics and Smart environments

Table of contents

Team DEFROST

Creation of the Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

2.3.3. - Real-time systems
3.1.1. - Modeling, representation
5.10. - Robotics
5.10.1. - Design
5.10.3. - Planning
5.10.4. - Robot control
5.10.5. - Robot interaction (with the environment, humans, other robots)
5.10.8. - Cognitive robotics and systems
6.2.1. - Numerical analysis of PDE and ODE
6.2.6. - Optimization
6.4.3. - Observability and Controlability
6.4.4. - Stability and Stabilization
7.1. - Parallel and distributed algorithms

Other Research Topics and Application Domains:

2.5.1. - Sensorimotor disabilities
2.7. - Medical devices
2.7.1. - Surgical devices
5.1. - Factory of the future
5.6. - Robotic systems
5.7. - 3D printing
9.2. - Art
9.2.2. - Cinema, Television
9.2.4. - Theater

1. Members

Research Scientists

Christian Duriez [Team leader, Inria, Senior Researcher, HDR]
Olivier Goury [Inria, Researcher]

Faculty Members

Jeremie Dequidt [Univ. Lille I, Associate Professor]
Alexandre Kruszewski [Ecole Centrale de Lille, Associate Professor]

Technical Staff

Eulalie Coevoet [Inria]
Mario Sanz Lopez [Inria]
Damien Marchal [CNRS, Research Engineer]
Bruno Carrez [Inria, Research Engineer]

PhD Students

Frederick Largilliere [Univ. Lille III]
Thor Morales , Bieze [Univ. Lille I]
Maxime Thieffry [Univ. Valenciennes]
Zhongkai Zhang [Inria]

Visiting Scientists

Alejandro Rodriguez Aguilera [PhD Student, Visiting Scientist, from Mar 2016 until Jun 2016]
Ugo Chouinard [Inria, Polytechnique Montreal, PhD Student, from Sep 2016 until Dec 2016]

Administrative Assistant

Anne Rejl [Inria]

Other

Piyush Jain [Inria, from Apr 2016 until Sep 2016]

2. Overall Objectives

2.1. Overall Objectives

The DEFROST team aims to address the open problem of control and modelling methods for deformable robots by answering the following challenges:

- Providing numerical methods and software support to reach the real- time constraint needed by robotic systems: the numerical solutions for the differential equations governing the deformation generate tens of thousands degrees of freedom, which is three orders of magnitude of what is frequently considered in classical methods of robotic modelling and control.
- Integrating deformation models in the control methods of soft robot: In soft-robotics, sensing, actuation and motion are coupled by the deformations. Deformable models must be placed at the heart of the control algorithm design.
- Investigating predictable interaction models with soft-tissues and pa- rameter estimation by visual feedback from medical imaging: On the contrary to many cases in surgical robotics, the contact of the soft robot with the anatomy is permitted and it creates additional deformations on the robot.

3. Research Program

3.1. Introduction

Our research crosses different disciplines: numerical mechanics, control design, robotics, optimisation methods, clinical applications. Our organisation aims at facilitating the team work and cross- fertilisation of research results in the group. We have three objectives (1, 2 and 3) that correspond to the main scientific challenges. In addition, we have two transversal objectives that are also highly challenging: the development of a high performance software support for the project (objective 4) and the validation tools and protocols for the models and methods (objective 5).

3.2. Objective 1: Accurate model of soft robot deformation computed in finite time

The objective is to find concrete numerical solutions to the challenge of modelling soft robots with strong real-time constraints. To solve continuum mechanics equations, we will start our research with real-time FEM or equivalent methods that were developed for soft-tissue simulation. We will extend the functionalities to account for the needs of a soft-robotic system:

- Coupling with other physical phenomenons that govern the activity of sensors and actuators (hydraulic, pneumatic, electro-active polymers, shape-memory alloys...).
- Fulfill the new computational time constraints (harder than surgical simulation for training) and find better tradeoff between cost and precision of numerical solvers using reduced-order modelling techniques with error control.
- Exploring interactive and semi-automatic optimisation methods for design based on obtained solution for fast computation on soft robot models.

3.3. Objective 2: Model based control of soft robot behavior

The focus of this objective is on obtaining a generic methodology for soft robot feedback control. Several steps are needed to design a model based control from FEM approach:

- The fundamental question of the kinematic link between actuators, sensors, effectors and contacts us- ing the most reduced mathematical space must be carefully addressed. We need to find efficient algorithms for real-time projection of non-linear FEM models in order to pose the control problem using the only relevant parameters of the motion control.
- Intuitive remote control is obtained when the user directly controls the effector motion. To add this functionality, we need to obtain real-time inverse models of the soft robots by optimisation. Several criteria will be combined in this optimisation: effector motion control, structural stiffness of the robot, reduce intensity of the contact with the environment...
- Investigating closed-loop approaches using sensor feedback: as sensors cannot monitor all points of the deformable structure, the information provided will only be partial. We will need additional algorithms based on the FEM model to obtain the best possible treatment of the information. The final ob- jective of these models and algorithms is to have robust and efficient feedback control strategies for soft robots. One of the main challenge here is to ensure / prove stability in closed-loop.

3.4. Objective 3: Modeling the interaction with a complex environment

Even if the inherent mechanical compliance of soft robots makes them more safe, robust and particularly adapted to interaction with frag- ile environments, the contact forces need to be controlled by:

- Setting up real-time modelling and the control methods needed to pilot the forces that the robot imposes on its environment and to control the robot deformations imposed by its environment. Note that if an operative task requires to apply forces on the surrounding structures, the robot must be anchored to other structures or structurally rigidified.
- Providing mechanics models of the environment that include the uncertainties on the geometry and on the mechanical properties, and are capable of being readjusted in real-time.
- Using the visual feedback of the robot behavior to adapt dynamically the models. The observation provided in the image coupled with an inverse accurate model of the robot could transform the soft robot into sensor: as the robot deforms with the contact of the surroundings, we could retrieve some missing parameters of the environment by a smart monitoring of the robot deformations.

3.5. Objective 4: Soft Robotic Software

Expected research results of this project are numerical methods and algorithms that require high-performance computing and suitability with robotic applications. There is no existing software support for such development. We propose to develop our own software, in a suite split into three applications:

- The first one will facilitate the design of deformable robots by an easy passage from CAD software (for the design of the robot) to the FEM based simulation
- The second one is an anticipative clinical simulator. The aim is to co-design the robotic assistance with the physicians, thanks to a realistic simulation of the procedure or the robotic assistance. This will facilitate the work of reflection on new clinical approaches prior any manufacturing
- The third one is the control design software. It will provide the real-time solutions for soft robot control developed in the project.

3.6. Objective 5: Validation and application demonstrations

The implementation of experimental valida- tion is a key challenge for the project. On one side, we need to validate the model and control algorithms using concrete test case example in order to improve the modelling and to demonstrate the concrete feasibility of our methods. On the other side, concrete applications will also feed the reflexions on the objec- tives of the scientific program.

We will build our own experimental soft robots for the validation of objective 2 and 3 when there is no existing « turn-key » solution. Designing and making our own soft robots, even if only for validation, will help the setting-up of adequate models.

For the validation of objective 4, we will develop « anatomical soft robot »: soft robot with the shape of organs, equipped with sensors (to measure the contact forces) and actuators (to be able to stiffen the walls and recreate natural motion of soft-tissues). We will progressively increase the level of realism of this novel validation set-up to come closer to the anatomical properties.

4. Application Domains

4.1. Industry

Robotics in the manufacturing industry is already highly diffused and is one of the ways put forward to maintain the level of competitiveness of companies based in France and to avoid relocation in cheap labor countries. Yet, in France, it is considered that the level of robotization is insufficient compared to Germany for instance. One of the challenge is the high investment cost for buying robotic arms. In the recent years, it has led the development of « generic » and « flexible » (but rigid) robotic solution that can be produced in series. But their applicability to specific tasks is still challenging or too costly. With the development of 3D printing, we can imagine the development of a complete opposite strategy: a « task-specific » design of robots. Given a task that need to be performed by a deformable robot: we would optimize the shape of its structure to create the set of desired motion . An second important aspect is the reduction of the manufacturing cost: It is often anticipated that the cost of deformable robots will be low compared to classical rigid robotics. The robot could be built on one piece using rapid prototyping or 3D printers and be more adapted for collaborative work with operators. This remains to be « proved », but it could open new perspectives in robotic applications. A last remarkable property of soft robots is their adaptability to fragile or tortuous environment. For some particular industry (chemistry, food industry...) this could also be an advantage compared to existing rigid solutions. For instance, the german company http://www.festo.com/Festo, key player in the industrial robots field, is experiencing with deformable trunk robot and we are working on their accurate control.

4.2. Personal and service robotics

The personal and service robotics are considered as an important source of economic expansion in the coming years. The potential applications are numerous and particularly include the challenge of finding robotic solutions for active and healthy aging at home. We plan to develop functional orthosis for which it is better not to have a rigid exoskeleton that are particularly not comfortable. These orthosis will be ideally personalised for each patient and built using rapid prototyping. On this topic, the place of our team will be to provide algorithms for controlling the robots. We will find some partners to build these robots that would fall in the category of « wearable robots ». With this thematic we also connect with a strong pole of excellence of the region on intelligent textile Up-Tex.

4.3. Entertainment industry and arts

Robots have a long history with entertainment and arts where animatronics have been used since years for cinematographic shootings, theater, amusement parc and performing arts. These animatronics are either radio-controlled by a team of professionals or using recorded movements. Our FEM-inversed approach to control soft robots may simplify animatronic control and thus impact this field. We are currently working on implementing demonstration of a deformable animatronic puppets in which motion tracking systems are used and the gestures and movements directly control the puppet. We are also collaborating with the art school Le Fresnoy based at Tourcoing, in particular with the artist Jonathan Pepe (see figure 1).

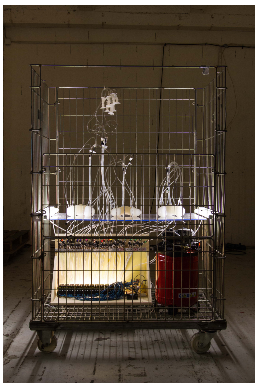


Figure 1. Our team has worked with the artist Jonathan Pepe on this art work that will be presented at the museum Le palais de Tokyo in 2017.

5. Highlights of the Year

5.1. Highlights of the Year

New Research scientist

Olivier Goury was selected to join the team as new Inria research scientist.

Robosoft Grand Challenge

The team participated in the Robosoft Week in Livorno, with a workshop on simulation of soft robots held by Christian Duriez, Thor Bieze and Eulalie Coevoet. In addition, 2 prototypes were presented to the Robosoft Grand Challenge, reaching the 4th place of the competition.

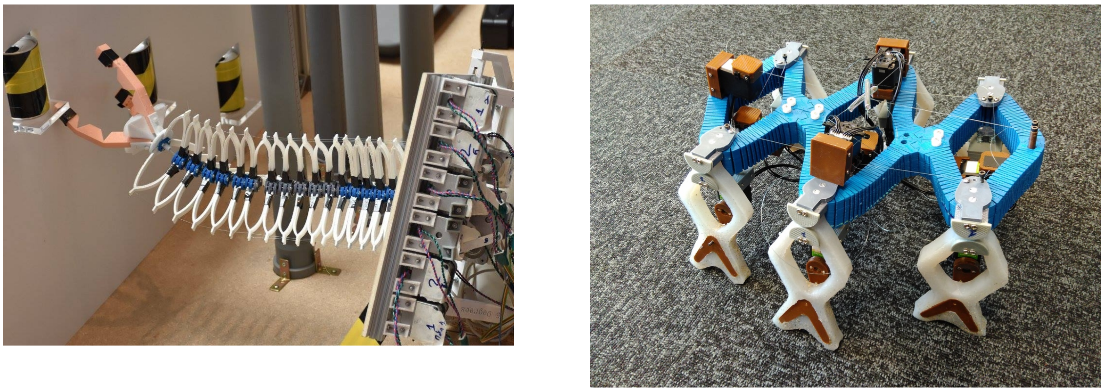


Figure 2. (a) EchelonIII (b) SOFIA

ERC evaluation grade A

The project COMOROS submitted for ERC Consolidator "fully met the ERC's excellence criterion" and evaluated as grade A. Unfortunately, it could not be funded, given the available budgetary resources of ERC for the call. But the region Haut-De-France should be able to finance a part of the project during the 3 coming years thanks to the FEDER funds.

6. New Software and Platforms

6.1. SOFA

Simulation Open Framework Architecture
KEYWORDS: Real time - Multi-physics simulation - Medical applications
FUNCTIONAL DESCRIPTION SOFA is an Open Source framework primarily targeted at real-time simulation, with an emphasis on medical simulation. It is mostly intended for the research community to help develop new algorithms, but can also be used as an efficient prototyping tool. Based on an advanced software architecture, it allows : the creation of complex and evolving simulations by combining new algorithms with algorithms already included in SOFA, the modification of most parameters of the simulation (deformable behavior, surface representation, solver, constraints, collision algorithm, etc.) by simply editing an XML file, the building of complex models from simpler ones using a scene-graph description, the efficient simulation of the dynamics of interacting objects using abstract equation solvers, the reuse and easy comparison of a variety of available methods.
Since 2016, SOFA development and maintenance is now coordinated by the SOFA Consortium.
DEFROST is an active member of the consortium steering committee; beside his mission of Inria continuous integration support team coordinator, Bruno Carrez is in charge of the continuous integration setup of the SOFA consortium.

- Participants: Christian Duriez, Jeremie Dequidt, Bruno Carrez, Damien Marchal, Eulalie Coevoet, Frederick Largilliere
- Partner: Sofa consortium, projet-team Mimesis, projet-team Imagine, projet-team Asclepios, InSimo, Anatoscope
- Contact: Hugo Talbot
- URL: http://www.sofa-framework.org

6.2. Soft-robot plugin for Sofa

Soft-robot plugin for Sofa
KEYWORDS: Simulation - Soft-Robot - Inverse models - Finite Element Method - Quadratic Programmings
FUNCTIONAL DESCRIPTION The soft-robot plugin consists in a new framework to simulate and control soft robots. This framework is based on a mechanical modeling of the robot elements in Sofa combined with fast real-time direct/inverse FEM solvers.
The keypoint of the approach implemented is that the same modeling is used for interactive simulation of its behavior and interactive control of the fabricated robots. This plugin was developped during the ADT project SORBET that ended in 09/2016.

- Participants: Eulalie Coevoet, Olivier Goury, Frederick Largilliere, Bruno Carrez, Damien Marchal, Jérémie Dequidt and Christian Duriez
- Contact: Eulalie Coevoet and Christian Duriez
- URL: https://project.inria.fr/softrobot/

7. New Results

7.1. Cochlear Implants

Publication at MICCAI 2016 (Medical Image Computing and Computer Assisted Intervention conference): **Numerical Simulation of Cochlear-Implant Surgery: Towards Patient-Specific Planning**, *Olivier Goury, Yann Nguyen, Renato Torres, Jeremie Dequidt, Christian Duriez*. **Abstract.** During Cochlear Implant Surgery, the right placement of the implant and the minimization of the surgical trauma to the inner ear are an important issue with recurrent fails. In this study, we reproduced, using simulation, the mechanical insertion of the implant during the surgery. This simulation allows to have a better understanding of the failing cases: excessive contact force, buckling of the implant inside and outside the cochlea. Moreover, using a patient-specific geometric model of the cochlea in the simulation, we show that the insertion angle is a clinical parameter that has an influence on the forces endured by both the cochlea walls and the basilar membrane, and hence to post-operative trauma. The paper presents the mechanical models used for the implant, for the basilar membrane and the boundary conditions (contact, friction, insertion etc...) and discuss the obtained results in the perspective of using the simulation for planning and robotization of the implant insertion.

https://hal.archives-ouvertes.fr/hal-01370185

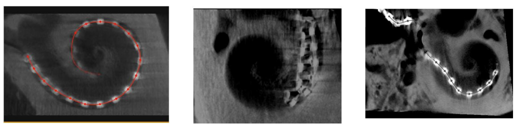


Figure 3. Three outcomes of implant insertion (from left to right): successful insertion; failed insertion (Folding tip); incomplete insertion

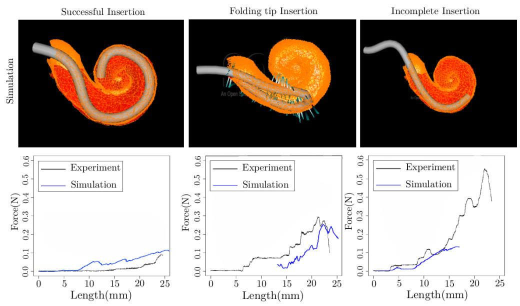


Figure 4. Reproduction of real insertion cases with the simulation

7.2. Physics based model of soft-robots

Book chapter in Soft Robotics: Trends, Applications and Challenges, Springer, 2016 **Soft Robot Modeling, Simulation and Control in Real-Time**, *Christian Duriez and Thor Bieze*. https://hal.inria.fr/hal-01410293. We were asked to write a chapter in this book on Soft Robotics. Our chapter presents new real-time and physics-based modeling methods dedicated to deformable soft robots. In our approach, continuum mechanics provides the partial derivative equations that govern the deformations, and Finite Element Method (FEM) is used to compute numerical solutions adapted to the robot. A formulation based on Lagrange Multipliers is presented to model the behavior of the actuators as well as the contact with the environment. Direct and inverse kinematic models are also obtained for real-time control. Some experiments and numerical results are presented.

7.3. Kinematic Modeling and control of soft robots

Publication at IROS 2016 : **Kinematic Modeling and Observer Based Control of Soft Robot using Real-Time Finite Element Method**, *Zhongkai Zhang, Jeremie Dequidt, Alexandre Kruszewski, Frederick Largilliere, Christian Duriez.*. **Abstract.**This paper aims at providing a novel approach to modeling and controlling soft robots. Based on real-time Finite Element Method (FEM), we obtain a globally defined discrete-time kinematic model in the workspace of soft robots. From the kinematic equations, we deduce the soft-robot Jacobian matrix and discuss the conditions to avoid singular configurations. Then, we propose a novel observer based control methodology where the observer is built by Finite Element Model in this paper to deal with the control problem of soft robots. A closed-loop controller for position control of soft robot is designed based on the discrete-time model with feedback signal being extracted by means of visual servoing. Finally, experimental results on a parallel soft robot show the efficiency and performance of our proposed controller. https://hal.inria.fr/hal-01370347

7.4. Stiffness rendering

Publication at IROS 2016 : **Stiffness rendering on soft tangible devices controlled through inverse FEM simulation**, *Frederick Largilliere, Eulalie Coevoet, Mario Sanz-Lopez, Laurent Grisoni, Christian Duriez.* **Abstract.**Haptic rendering of soft bodies is essential in medical simulations of procedures such as surgery or palpation. The most commonly used approach is to recreate the sense of touch using a specific design and control of a robotic arm. In this paper, we propose a new approach, based on soft-robotics technology. We create a tangible deformable device that allows users to " touch " soft tissues and perceive mechanical material properties, in a realistic manner. The device is able to dynamically provide user touch with different stiffness perceptions, thanks to actuators placed at the boundaries. We introduce a control algorithm, based on inverse Finite Element Analysis, which controls the actuators in order to recreate a desired stiffness that corresponds to the contact with soft tissues in the virtual environment. The approach uses antagonistic actuation principle to create a wide range of stiffness. We validate our algorithm and demonstrate the method using prototypes based on simple mechanisms. https://hal.inria.fr/hal-01386787

7.5. Framework for soft robot simulation

Publication at SIMPAR 2016 : **Framework for online simulation of soft robots with optimization-based inverse model**, *C. Duriez, E. Coevoet, F. Largilliere, T. Morales-Bieze, Z. Zhang, M. Sanz-Lopez, B. Carrez, D. Marchal, O. Goury, J. Dequidt*. **Abstract.**Soft robotics is an emerging field of robotics which requires computer-aided tools to simulate soft robots and provide models for their control. Until now, no unified software framework covering the different aspects exists. In this paper, we present such a framework from its theoretical foundations up to its implementation on top of Sofa, an open-source framework for deformable online simulation. The framework relies on continuum mechanics for modeling the robotic parts and boundary conditions like actuators or contacts using a unified representation based on Lagrange multipliers. It enables the digital robot to be simulated in its environment using a direct model. The model can also be inverted online using an optimization-based method which allows to control the physical robots in the task space. To demonstrate the effectiveness of the approach, we present various soft robots scenarios including ones where the robot is interacting with its environment. https://hal.inria.fr/hal-01425349

7.6. Closed-loop control

Closed-loop control based on dynamic models of soft robots. Model-order reduction provides a system of achievable size to apply traditional control science techniques. During the internship of Maxime Thieffry, we obtain the first results in that direction that will be extended during a PhD thesis.

8. Bilateral Contracts and Grants with Industry

8.1. A.I. Mergence

A.I. Mergence is a startup company based in Paris. The transfer contract was about building a soft robot prototype. The aim of the demonstration was to show that we can improve the appearance and user interaction. They have a usage of our license for 12 months. Amount of the contract: 1500 euros.

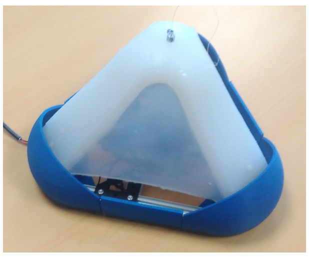


Figure 5. Prototype for A.I. Mergence

8.2. TruPhysics

TruPhysics is a german startup, using SOFA for the simulation of industrial robots. We did an expertise and research contract on modeling grasping tasks in SOFA with a deformable gripper. Amount of the contract: 7940 euros.

8.3. InSimo

InSimo is a French startup, based in Strasbourg, that was created by members of the team in 2013. The goal of InSimo is to create a new generation of surgical simulators with high quality biomechanics. We have signed a contract to work on the simulation of suture during the years 2016-2017. Amount of the contract: 33000 euros.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. Inserm

Olivier Goury was hired as a postdoctoral researcher by the "Réhabilitation chirurgicale mini-invasive et robotisée de l'audition" to collaborate with the DEFROST team on the simulation of Cochlear Implant surgery. The contract stopped since Olivier has been recruited as a Research scientist. The collaboration with Inserm will be continued with the hiring of Piyush Jain as an engineer.

9.1.2. ANR

- **Sorcery** The goal of this project was to work on the modeling, simulation and control of soft surgical robot with a particular focus in cochlear implantology. A very good consortium was built around the project that went to phase 2 in the ANR project. Unfortunately, the project has not been funded.
- **IDeaS**, Image-Driven Simulation, Jeremie Dequidt, Magrit, MIMESIS and Nancy Hospital, 42 months,: this is a project targeted at per-operative guidance for interventional radiology procedures. Our main goal is to provide effective solutions for the two main drawbacks of interventional radiology procedures, namely: reduce radiation exposure and provide a fully 3D and interactive visual feedback during the procedure. To do so, our project relies on an original combination of computer vision algorithms and interactive physics-based medical simulation.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

Program: FET Open

Project acronym: RoboSoft

Project title: Coordination Action for Soft Robotics

Duration: 2014-2016

Coordinator: Cecilia Laschi (Scuola Superiore Sant'Anna)

9.3. International Initiatives

9.3.1. Declared Inria International Partners

We have a collaboration with King's College (Profs Kaspar Althoefer and Hongbin Liu) on soft hydraulic robots with the support of the program North European associate team of the center Inria Lille North Europe (2014-2016)

We have started a collaboration with the Université Libre de Bruxelles (Profs Denis Terwagne, Serge Massar, Marc Haelterman and Guillaume Tillema) on the use of soft robot simulation to build control strategies based on artificial intelligence algorithms (2016-2018)

9.3.2. Informal International Partners

This section includes some recent collaboration. We have initiated research work with Prof. Bordas at the University of Luxembourg on Model reduction with contacts. We are also working with Adrien Escande and with Prof. Yoshida, at AIST Japan, on the simulation of deformable objects in contact and with Prof Miguel Otaduy at URJC Madrid on human hand grasping and manipulation (Conference paper in 2014 and journal paper in 2015).

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Ugo Chouinard was a research intern in the Defrost team from September to December 2016. He is a PhD candidate in mechanical engineering at Polytechique Montréal (Canada) and obtained the Mitacs-Globalink Research Award that allowed him to join our team for a few months. He investigated the effect of design change on the compliance of deformable manipulators. The result of his research will help to better understand and design soft robotic manipulators. Indeed, with the research he carried out, it will be easier to design robots that meet the design specifications. Furthermore, his internship might lead to further collaboration with Polytechnique Montréal for the design of soft robotics systems.
- Alejandro Rodriguez Aguilera from the University of Granada, stayed from March to June 2016. His works on GPU computing allowed him to develop a parallelized hydraulics systems simulation and integrate it into the SOFA Framework.

9.4.1.1. Internships

- Valentin Owczarek was a research intern in the Defrost team from March to September 2016. He worked on using genetics algorithm to generate task specific soft-robot designs.
- During the internship of Piyush Jain (India) from April 2016 to September 2016, it was observed that it is possible to create the concept of self-contained pneumatic actuation for soft robots without the need for an external pneumatic supply.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Conference Program Committees

- Jeremie Dequidt is member of the CP committee of the AFIG (Association Française d'Informatique Graphique) conference and VRIPHYS conference
- Christian Duriez is member of the CP committee of Eurohaptics

10.1.1.2. Reviewer

- Jeremie Dequidt has been reviewer for International Journal of Computer Assisted Radiology and Surgery (IJCARS) and AFIG 2016
- Damien Marchal has been reviewer for IEEE 3D User Interface 2016 (3DUI 2016) and EuroHaptic 2016.
- Christian Duriez has been reviewer for the conferences IEEE VR 2016 and ICRA 2017.

10.1.2. Journal

10.1.2.1. Member of the Editorial Boards

- Christian Duriez is member of the editorial board of IEEE Transaction on Haptics

10.1.2.2. Reviewer - Reviewing Activities

- Christian Duriez has been reviewer for the journals ACM Transaction on Graphics, Computer and Graphics, IEEE Transactions on Robotics, Soft Robotic Journal (SORO).

10.1.3. Invited Talks

- Christian Duriez and Thor Bieze were invited for a talk at the Spring School on Soft-robotics at Livorno (April 2016)
- Christian Duriez was invited for a talk at the conference POSS in Paris (november 2016).

10.1.4. Scientific Expertise

Christian Duriez is a scientific advisor of InSimo (Inria Spin-off)

10.1.5. Research Administration

Christian Duriez is vice-responsible of Research Jobs Committee at Inria Lille.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence : Jeremie Dequidt, Programming, 54h, L3, University of Lille/Polytech Lille, France

Licence : Jeremie Dequidt, Advance Programming, 16h, L3, University of Lille/Polytech Lille, France

Licence : Jeremie Dequidt, Databases, 12h, L3, University of Lille/Polytech Lille, France

Licence : Frederick Largilliere, Computer science fundamentals, 48h, L1, University of Lille 3, France

Licence : Frederick Largilliere, Certificate of computer science and internet knowledge, 27.5h, L2 - L3, University of Lille 3, France

Master : Jeremie Dequidt, Theoretical Computer Science, 24h, M1, University of Lille/Polytech Lille, France

Master : Jeremie Dequidt, Surgical Simulation, 12h, M2, University of Lille/Polytech Lille, France

Master : Thor Morales Bieze, Bio-inspired robotics, 20h, M2, University of Lille/Polytech Lille, France

Master : Damien Marchal, Reality-Virtual et Interaction, 20h, M2, University of Lille, France

Master : Christian Duriez, Interactive hysics-based enginesh, M2, University of Lille, France

10.2.2. Supervision

PhD in progress: Frederick Largilliere, Stiffness control of interfaces based on soft robots through numerical modelization, 01/11/2013, Laurent Grisoni, Christian Duriez

PhD in progress: Thor Morales Bieze, Design, modeling and control of soft, continuum manipulators, 10/10/2013, Rochdi Merzouki, Christian Duriez

PhD in progress: Zhongkai Zhang, Visual servoing control of soft robots based on real-time Finite Element Method, 01/10/2015, Jeremie Dequidt, Christian Duriez

PhD in progress: Maxime Thieffry, High velocity control of soft robots, 01/10/2016, Christian Duriez, Alexandre Kruszewski, Thierry Marie Guerra

10.2.3. Juries

- Christian Duriez was president of the committee of the PhD jury of José Dolz (University of Lille 2)

- Christian Duriez was reviewer of the PhDs of Johan Sarrazin (University of Grenoble) and Pierre-Luc Manteaux (University of Grenoble)
- Christian Duriez was examiner of the PhD of Zilong Shao (University of Lille)

10.3. Popularization

- Olivier Goury and Christian Duriez participated in "Fête de la science: Opération Chercheurs itinérants", which involves giving scientific lectures in middle and high schools.
- Christian Duriez and Mario Sanz Lopez participated in the "Bourse aux Technologies" at Bercy, Paris (22 March).
- Popularization session 13:45 on the topic of Arduino by Mario Sanz Lopez.
- Popularization session 13:45 on the topic of Software development good practices by Bruno Carrez.
- Damien Marchal participated in "Chti'Code". Chti'Code is an action of the University of Lille 1, which aims to promoting computer education in primary schools. Damien Marchal intervened with computer science student in 4 CM2 classes for a total of 20h.
- Jeremie Dequidt was jury member of the second robotic competition SI / @ Baggio.

11. Bibliography

Major publications by the team in recent years

[1] J. ALLARD, F. FAURE, H. COURTECUISSE, F. FALIPOU, C. DURIEZ, P. KRY.*Volume Contact Constraints at Arbitrary Resolution*, in "ACM Transactions on Graphics", August 2010, vol. 29, n^o 3, https://hal.inria.fr/inria-00502446.

[2] J. BOSMAN, T. M. BIEZE, O. LAKHAL, M. SANZ, R. MERZOUKI, C. DURIEZ.*Domain decomposition approach for FEM quasistatic modeling and control of Continuum Robots with rigid vertebras*, in "2015 IEEE International Conference on Robotics and Automation (ICRA)", May 2015, p. 4373-4378 [*DOI :* 10.1109/ICRA.2015.7139803].

[3] E. COEVOET, N. REYNAERT, E. LARTIGAU, L. SCHIAPPACASSE, J. DEQUIDT, C. DURIEZ.*Introducing interactive inverse FEM simulation and its application for adaptive radiotherapy*, in "MICCAI - 17th International Conference on Medical Image Computing and Computer-Assisted Intervention", Boston, United States, September 2014, https://hal.inria.fr/hal-01059667.

[4] H. COURTECUISSE, J. ALLARD, P. KERFRIDEN, S. P. BORDAS, S. COTIN, C. DURIEZ.*Real-time simulation of contact and cutting of heterogeneous soft-tissues*, in "Medical image analysis", 2014, vol. 18, n^o 2, p. 394–410.

[5] C. DURIEZ, E. COEVOET, F. LARGILLIERE, T. M. BIEZE, Z. ZHANG, M. SANZ-LOPEZ, B. CARREZ, D. MARCHAL, O. GOURY, J. DEQUIDT.*Framework for online simulation of soft robots with optimization-based inverse model*, in "SIMPAR: IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots", San Francisco, United States, Proceedings of SIMPAR 2016 conference, December 2016, https://hal.inria.fr/hal-01425349.

[6] C. DURIEZ, F. DUBOIS, A. KHEDDAR, C. ANDRIOT.*Realistic haptic rendering of interacting deformable objects in virtual environments*, in "IEEE Transactions on Visualization and Computer Graphics", January 2006, vol. 12, n^o 1, p. 36-47 [*DOI :* 10.1109/TVCG.2006.13].

[7] C. DURIEZ.*Control of elastic soft robots based on real-time finite element method*, in "Robotics and Automation (ICRA), 2013 IEEE International Conference on", IEEE, 2013, p. 3982–3987.

[8] C. DURIEZ.*Real-time haptic simulation of medical procedures involving deformations and device-tissue interactions*, Université des Sciences et Technologie de Lille - Lille I, February 2013, https://tel.archives-ouvertes.fr/tel-00785118.

[9] C. DURIEZ.*Method for Controlling a Deformable Robot, and Related Module and Computer Program*, December 2015, n^o US 20150360367, http://www.freepatentsonline.com/y2015/0360367.html.

[10] N. HAOUCHINE, J. DEQUIDT, I. PETERLIK, E. KERRIEN, M.-O. BERGER, S. COTIN.*Image-guided Simulation of Heterogeneous Tissue Deformation For Augmented Reality during Hepatic Surgery*, in "ISMAR - IEEE International Symposium on Mixed and Augmented Reality 2013", Adelaide, Australia, October 2013, https://hal.inria.fr/hal-00842855.

[11] F. LARGILLIERE, V. VERONA, E. COEVOET, M. SANZ-LOPEZ, J. DEQUIDT, C. DURIEZ.*Real-time control of soft-robots using asynchronous finite element modeling*, in "2015 IEEE International Conference on Robotics and Automation (ICRA)", May 2015, p. 2550-2555 [*DOI :* 10.1109/ICRA.2015.7139541].

[12] F. LARGILLIÈRE, E. COEVOET, M. SANZ-LOPEZ, L. GRISONI, C. DURIEZ.*Stiffness rendering on soft tangible devices controlled through inverse FEM simulation*, in "International Conference on Intelligent Robots and Systems - IROS 2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01386787.

[13] A. TALVAS, M. MARCHAL, C. DURIEZ, M. A. OTADUY.*Aggregate Constraints for Virtual Manipulation with Soft Fingers*, in "IEEE Transactions on Visualization and Computer Graphics", January 2015, vol. 21, n^o 4, 9, https://hal.inria.fr/hal-01246562.

[14] Z. ZHANG, J. DEQUIDT, A. KRUSZEWSKI, F. LARGILLIERE, C. DURIEZ.*Kinematic Modeling and Observer Based Control of Soft Robot using Real-Time Finite Element Method*, in "IROS2016 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01370347.

Publications of the year

Articles in International Peer-Reviewed Journal

[15] E. KERRIEN, A. YUREIDINI, J. DEQUIDT, C. DURIEZ, R. ANXIONNAT, S. COTIN.*Blood vessel modeling for interactive simulation of interventional neuroradiology procedures*, in "Medical Image Analysis", January 2017, vol. 35, p. 685 - 698 [*DOI :* 10.1016/J.MEDIA.2016.10.003], https://hal.inria.fr/hal-01390923.

[16] S. MAALEJ, A. KRUSZEWSKI, L. BELKOURA.*Robust Control for Continuous LPV System with Restricted-Model-Based Control*, in "Circuits, Systems, and Signal Processing", September 2016 [*DOI :* 10.1007/s00034-016-0404-6], https://hal.archives-ouvertes.fr/hal-01410300.

[17] R. MÁRQUEZ, T. M. GUERRA, M. BERNAL, A. KRUSZEWSKI.*A non-quadratic Lyapunov functional for $H\infty$ control of nonlinear systems via Takagi–Sugeno models*, in "Journal of The Franklin Institute", March 2016, vol. 353, n^o 4, p. 781 - 796 [*DOI :* 10.1016/J.JFRANKLIN.2016.01.004], https://hal.archives-ouvertes.fr/hal-01410230.

[18] R. MÁRQUEZ, T. M. GUERRA, M. BERNAL, A. KRUSZEWSKI.*Asymptotically necessary and sufficient conditions for Takagi–Sugeno models using generalized non-quadratic parameter-dependent controller design*, in "Fuzzy Sets and Systems", January 2016, vol. 306, p. 48 - 62 [*DOI :* 10.1016/J.FSS.2015.12.012], https://hal.archives-ouvertes.fr/hal-01410325.

[19] H. TALBOT, F. SPADONI, C. DURIEZ, M. SERMESANT, S. COTIN, H. DELINGETTE.*Interactive Training System for Interventional Electrocardiology Procedures*, in "Lecture notes in computer science", 2016, vol. 8789, p. 11-19 [*DOI :* 10.1007/978-3-319-12057-7_2], https://hal.inria.fr/hal-01338346.

International Conferences with Proceedings

[20] C. DURIEZ, E. COEVOET, F. LARGILLIERE, T. M. BIEZE, Z. ZHANG, M. SANZ-LOPEZ, B. CARREZ, D. MARCHAL, O. GOURY, J. DEQUIDT.*Framework for online simulation of soft robots with optimization-based inverse model*, in "SIMPAR: IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots", San Francisco, United States, Proceedings of SIMPAR 2016 conference, December 2016, https://hal.inria.fr/hal-01425349.

[21] O. GOURY, Y. NGUYEN, R. TORRES, J. DEQUIDT, C. DURIEZ.*Numerical Simulation of Cochlear-Implant Surgery: Towards Patient-Specific Planning*, in "19th International Conference on Medical Image Computing & Computer Assisted Intervention (MICCAI 2016)", Athènes, Greece, Lecture Notes in Computer Science, October 2016, vol. 9900, p. 500-507, https://hal.archives-ouvertes.fr/hal-01370185.

[22] F. LARGILLIÈRE, E. COEVOET, M. SANZ-LOPEZ, L. GRISONI, C. DURIEZ.*Stiffness rendering on soft tangible devices controlled through inverse FEM simulation*, in "International Conference on Intelligent Robots and Systems - IROS 2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01386787.

[23] Z. ZHANG, J. DEQUIDT, A. KRUSZEWSKI, F. LARGILLIERE, C. DURIEZ.*Kinematic Modeling and Observer Based Control of Soft Robot using Real-Time Finite Element Method*, in "IROS2016 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01370347.

Scientific Books (or Scientific Book chapters)

[24] C. DURIEZ, T. MORALES BIEZE.*Soft robot modeling, simulation and control in real-time*, in "Soft Robotics: Trends, Applications and Challenges", September 2016 [*DOI :* 10.1007/978-3-319-46460-2_13], https://hal.inria.fr/hal-01410293.

Project-Team DOLPHIN

Parallel Cooperative Multi-criteria Optimization

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Optimization, machine learning and statistical methods

Table of contents

Project-Team DOLPHIN

Creation of the Project-Team: 2005 May 12, updated into Team: 2017 January 01

Keywords:

Computer Science and Digital Science:

1.1.4. - High performance computing
1.1.5. - Exascale
3.1.4. - Uncertain data
6. - Modeling, simulation and control
7.1. - Parallel and distributed algorithms
7.3. - Optimization

Other Research Topics and Application Domains:

1. - Life sciences
2.7. - Medical devices
4. - Energy
7. - Transport and logistics
8.1.1. - Energy for smart buildings

1. Members

Research Scientists

Dimo Brockhoff [Inria, Researcher, until Sep 2016]
Fanny Dufosse [Inria, Researcher]

Faculty Members

El-Ghazali Talbi [Team leader, Univ. Lille I, Faculty Member, HDR]
Bilel Derbel [Univ. Lille I, Faculty Member]
Clarisse Dhaenens [Univ. Lille I, Faculty Member, HDR]
Marie-Eleonore Kessaci [Univ. Lille I, Faculty Member]
Arnaud Liefooghe [Univ. Lille I, Faculty Member]
Nouredine Melab [Univ. Lille I, Faculty Member, HDR]
Laetitia Vermeulen [Univ. Lille I, Faculty Member, HDR]

Technical Staff

Dejan Tušar [Inria, until Jul 2016]

PhD Students

Oumayma Bahri [Univ. Lille I]
Aymeric Blot [ENS Cachan]
Tiago Carneiro Pessoa [Universidade Federal Do Cearà, Brazil, until Aug 2016]
Sohrab Faramarzi Oghani [Inria]
Asmae Gannouni [Univ. Lille I]
Zineb Garroussi [EMI, University of Rabat, Morocco]
Jan Gmys [Université de Mons, Belgium]
Sophie Jacquin [Univ. Lille I, until Aug 2016]
Gauvain Marquet [Univ. Lille I]
Lucien Mousin [Min. Ens. Sup. Recherche]
Thanh-Do Tran [Inria CORDI, until Mar 2016]

Maxence Vandromme [ALICANTE, granted by CIFRE]

Post-Doctoral Fellows

Pornpimol Chaiwuttisak [Univ. Lille I, from Oct 2016]
Raca Todosijevic [Inria, from Feb 2016]
Tea Tušar [Inria, until Jul 2016]

Visiting Scientists

Oliver Cuate Gonzalez [CINVESTAV (Mexico), from Apr 2016 until Jun 2016]
Miyako Sagawa [Shinshu University (Japan), until Jun 2016]
Rachid Ellaia [EMI, University of Rabat, Morocco, until Nov 2016]
Lakhdar Loukil [Université d'Oran, Algeria, until Jan 2016]

Administrative Assistant

Julie Jonas [Inria]

Others

Nicolas Berveglieri [Univ. Lille I, BSc Student, from Apr 2016 until Jun 2016]
Mariette Biernacki [Inria, Polytech'Lille Internship, from Jun 2016 until Jul 2016]
Bastien Courtois [Univ. Lille I, Polytech'Lille Internship, from Sep 2016 until Nov 2016]
Christopher Lievin [Univ. Lille I, IUT Internship, from Apr 2016 until Jun 2016]
Alexis Pernet [Univ. Lille I, Master Lille 1 Internship, from Jun 2016 until Jul 2016]
Beatrice Verheyde [Univ. Lille I, Master Lille 1 Internship, until Jun 2016]

2. Overall Objectives

2.1. Presentation

The goal of the DOLPHIN [0] project is the modeling and resolution of large multi-criteria combinatorial problems using parallel and distributed hybrid techniques. We are interested in algorithms using Pareto approaches, which generate the whole Pareto set of a given Multi-Objective Problem (MOP). For this purpose, the research actions can be summarized as follows:

- **Modeling and Analysis of MOPs:** Solving Multi-Objective Problems requires an important analysis phase to find the best suitable method to solve it. This analysis deals with the modeling of the problem and the analysis of its structure.

  To propose efficient models for a Multi-Objective Optimization problem, an important aspect is to integrate all the constraints of the problem. Therefore an interesting preliminary approach is to develop efficient models for the problem in its mono-objective forms in order to be able to develop methods that are taking the characteristics of the studied problem into account.

  While studying the problem in its multi-objective form, the analysis of the structure is another interesting approach. The analysis of the structure of the Pareto front by means of different approaches (statistical indicators, meta-modeling, etc.) allows the design of efficient and robust hybrid optimization techniques. In general, the current theory does not allow the complete analysis of optimization algorithms. Several questions are unanswered: i) why is a given method efficient? ii) why are certain instances difficult to solve? Some work is needed to guide the user in the design of efficient methods.

  The NFL (No Free Lunch) theorem shows that two optimization methods have the same global performance on the whole set of uniform optimization problems. Then, it is crucial to make some hypotheses on the studied problem. This may be done in two steps:

  – analyzing the target problem to identify its landscape properties,
  – including this knowledge in the proposed optimization method.

[0] Discrete multi-objective Optimization for Large scale Problems with Hybrid dIstributed techNiques.

Our interest in this project is to answer these questions and remarks for the multi-objective case. Another point considered is the performance evaluation of multi-objective optimization methods. We are also working on approximation algorithms with performance guarantee and the convergence properties of stochastic algorithms.

- **Cooperation of optimization methods (metaheuristics and/or exact methods):**

  The hybridization of optimization methods allows the cooperation of complementary different methods. For instance, the cooperation between a metaheuristic and an exact method allows us to take advantage of the intensification process of an exact method in finding the best(s) solution(s) in a sub-space, and the diversification process of the metaheuristic in reducing the search space to explore.

  In this context, different types of cooperation may be proposed. These approaches are under study in the project and we are applying them to different generic MOPs (flow-shop scheduling problem, vehicle routing problem, covering tour problem, access network design, and the association rule problem in data mining).

- **Parallel optimization methods:** Parallel and distributed computing may be considered as a tool to speedup the search to solve large MOPs and/or to improve the robustness of a given method. Following this objective, we design and implement parallel metaheuristics (evolutionary algorithms, Tabu search approach) and parallel exact methods (branch and bound algorithm, branch and cut algorithm) for solving different large MOPs. Moreover, the joint use of parallelism and cooperation allows the improvement of the quality of the obtained solutions.

- **Framework for parallel and distributed hybrid metaheuristics:** Our team contributes to the development of an open source framework for metaheuristics, named ParadisEO (PARAllel and DIStributed Evolving Objects). Our contribution in this project is the extension of the EO (Evolving Objects) framework [0], which consists in: i) the generalization of the framework to single solution metaheuristics such as local search, tabu search and simulated annealing; ii) the design of metaheuristics for multi-objective optimization; iii) the design of hybrid methods; iv) the development of parallel and distributed models.

  In this project, our goal is the efficient design and implementation of this framework on different types of parallel and distributed hardware platforms:cluster of workstations (COW), networks of workstations (NOW) and GRID computing platforms, using the suited programming environments (MPI, Condor, Globus, PThreads). The coupling with well-known frameworks for exact methods (such as COIN) will also be considered. The exact methods for MOPs developed in this project will be integrated in those software frameworks.

  The experimentation of this framework by different users and applications outside the DOLPHIN project is considered. This is done in order to validate the design and the implementation issues of ParadisEO.

- **Validation:** the designed approaches are validated on generic and real-life MOPs, such as:
  1. Scheduling problems: Flow-shop scheduling problem.
  2. Routing problems: Vehicle routing problem (VRP), covering tour problem (CTP).
  3. mobile telecommunications: Design of mobile telecommunications networks (contract with France Telecom R&D) and design of access networks (contract with Mobinets).
  4. Genomics: Association rule discovery (data mining task) for mining genomic data, protein identification, docking and conformational sampling of molecules.
  5. Engineering design problems: Design of polymers.

[0] This framework was initially developed by Geneura TEAM (Spain), Inria (France), LIACS (Netherlands). http://paradiseo.gforge.inria.fr.

Some benchmarks and their associated optimal Pareto fronts or best known Pareto fronts have been defined and made available on the Web. We are also developing an open source software, named GUIMOO [0], which integrates different performance evaluation metrics and 2D/3D visualization tools of Pareto fronts.

3. Research Program

3.1. Hybrid multi-objective optimization methods

The success of metaheuristics is based on their ability to find efficient solutions in a reasonable time [54]. But with very large problems and/or multi-objective problems, efficiency of metaheuristics may be compromised. Hence, in this context it is necessary to integrate metaheuristics in more general schemes in order to develop even more efficient methods. For instance, this can be done by different strategies such as cooperation and parallelization.

The DOLPHIN project deals with *"a posteriori"* multi-objective optimization where the set of Pareto solutions (solutions of best compromise) have to be generated in order to give the decision maker the opportunity to choose the solution that interests him/her.

Population-based methods, such as evolutionary algorithms, are well fitted for multi-objective problems, as they work with a set of solutions [50], [53]. To be convinced one may refer to the list of references on Evolutionary Multi-objective Optimization maintained by Carlos A. Coello [0], which contains more than 5500 references. One of the objectives of the project is to propose advanced search mechanisms for intensification and diversification. These mechanisms have been designed in an adaptive manner, since their effectiveness is related to the landscape of the MOP and to the instance solved.

In order to assess the performances of the proposed mechanisms, we always proceed in two steps: first, we carry out experiments on academic problems, for which some best known results exist; second, we use real industrial problems to cope with large and complex MOPs. The lack of references in terms of optimal or best known Pareto set is a major problem. Therefore, the obtained results in this project and the test data sets will be available at the URL http://dolphin.lille.inria.fr/ at 'benchmark'.

3.1.1. Cooperation of metaheuristics

In order to benefit from the various advantages of the different metaheuristics, an interesting idea is to combine them. Indeed, the hybridization of metaheuristics allows the cooperation of methods having complementary behaviors. The efficiency and the robustness of such methods depend on the balance between the exploration of the whole search space and the exploitation of interesting areas.

Hybrid metaheuristics have received considerable interest these last years in the field of combinatorial optimization. A wide variety of hybrid approaches have been proposed in the literature and give very good results on numerous single objective optimization problems, which are either academic (traveling salesman problem, quadratic assignment problem, scheduling problem, etc) or real-world problems. This efficiency is generally due to the combinations of single-solution based methods (iterative local search, simulated annealing, tabu search, etc) with population-based methods (genetic algorithms, ants search, scatter search, etc). A taxonomy of hybridization mechanisms may be found in [56]. It proposes to decompose these mechanisms into four classes:

- *LRH class - Low-level Relay Hybrid*: This class contains algorithms in which a given metaheuristic is embedded into a single-solution metaheuristic. Few examples from the literature belong to this class.
- *LTH class - Low-level Teamwork Hybrid*: In this class, a metaheuristic is embedded into a population-based metaheuristic in order to exploit strengths of single-solution and population-based metaheuristics.

[0] Graphical User Interface for Multi-Objective Optimization (http://guimoo.gforge.inria.fr).
[0] http://delta.cs.cinvestav.mx/~ccoello/EMOO/EMOObib.html

- *HRH class - High-level Relay Hybrid*: Here, self contained metaheuristics are executed in a sequence. For instance, a population-based metaheuristic is executed to locate interesting regions and then a local search is performed to exploit these regions.
- *HTH class - High-level Teamwork Hybrid*: This scheme involves several self-contained algorithms performing a search in parallel and cooperating. An example will be the island model, based on GAs, where the population is partitioned into small subpopulations and a GA is executed per subpopulation. Some individuals can migrate between subpopulations.

Let us notice that, hybrid methods have been studied in the mono-criterion case, their application in the multi-objective context is not yet widely spread. The objective of the DOLPHIN project is to integrate specificities of multi-objective optimization into the definition of hybrid models.

3.1.2. Cooperation between metaheuristics and exact methods

Until now only few exact methods have been proposed to solve multi-objective problems. They are based either on a Branch-and-bound approach, on the algorithm $A^{\star}$, or on dynamic programming. However, these methods are limited to two objectives and, most of the time, cannot be used on a complete large scale problem. Therefore, sub search spaces have to be defined in order to use exact methods. Hence, in the same manner as hybridization of metaheuristics, the cooperation of metaheuristics and exact methods is also a main issue in this project. Indeed, it allows us to use the exploration capacity of metaheuristics, as well as the intensification ability of exact methods, which are able to find optimal solutions in a restricted search space. Sub search spaces have to be defined along the search. Such strategies can be found in the literature, but they are only applied to mono-objective academic problems.

We have extended the previous taxonomy for hybrid metaheuristics to the cooperation between exact methods and metaheuristics. Using this taxonomy, we are investigating cooperative multi-objective methods. In this context, several types of cooperations may be considered, according to the way the metaheuristic and the exact method cooperate. For instance, a metaheuristic can use an exact method for intensification or an exact method can use a metaheuristic to reduce the search space.

Moreover, a part of the DOLPHIN project deals with studying exact methods in the multi-objective context in order: i) to be able to solve small size problems and to validate proposed heuristic approaches; ii) to have more efficient/dedicated exact methods that can be hybridized with metaheuristics. In this context, the use of parallelism will push back limits of exact methods, which will be able to explore larger size search spaces [51].

3.1.3. Goals

Based on the previous works on multi-objective optimization, it appears that to improve metaheuristics, it becomes essential to integrate knowledge about the problem structure. This knowledge can be gained during the search. This would allow us to adapt operators which may be specific for multi-objective optimization or not. The goal here is to design auto-adaptive methods that are able to react to the problem structure. Moreover, regarding the hybridization and the cooperation aspects, the objectives of the DOLPHIN project are to deepen these studies as follows:

- *Design of metaheuristics for the multi-objective optimization:* To improve metaheuristics, it becomes essential to integrate knowledge about the problem structure, which we may get during the execution. This would allow us to adapt operators that may be specific for multi-objective optimization or not. The goal here is to design auto-adaptive methods that are able to react to the problem structure.
- *Design of cooperative metaheuristics:* Previous studies show the interest of hybridization for a global optimization and the importance of problem structure study for the design of efficient methods. It is now necessary to generalize hybridization of metaheuristics and to propose adaptive hybrid models that may evolve during the search while selecting the appropriate metaheuristic. Multi-objective aspects have to be introduced in order to cope with the specificities of multi-objective optimization.

- *Design of cooperative schemes between exact methods and metaheuristics:* Once the study on possible cooperation schemes is achieved, we will have to test and compare them in the multi-objective context.
- *Design and conception of parallel metaheuristics:* Our previous works on parallel metaheuristics allow us to speed up the resolution of large scale problems. It could be also interesting to study the robustness of the different parallel models (in particular in the multi-objective case) and to propose rules that determine, given a specific problem, which kind of parallelism to use. Of course these goals are not disjoined and it will be interesting to simultaneously use hybrid metaheuristics and exact methods. Moreover, those advanced mechanisms may require the use of parallel and distributed computing in order to easily make cooperating methods evolve simultaneously and to speed up the resolution of large scale problems.
- *Validation:* In order to validate the obtained results we always proceed in two phases: validation on academic problems, for which some best known results exist and use on real problems (industrial) to cope with problem size constraints.

  Moreover, those advanced mechanisms are to be used in order to integrate the distributed multi-objective aspects in the ParadisEO platform (see the paragraph on software platform).

3.2. Parallel multi-objective optimization: models and software frameworks

Parallel and distributed computing may be considered as a tool to speedup the search to solve large MOPs and to improve the robustness of a given method. Moreover, the joint use of parallelism and cooperation allows improvements on the quality of the obtained Pareto sets. Following this objective, we will design and implement parallel models for metaheuristics (evolutionary algorithms, tabu search approach) and exact methods (branch-and-bound algorithm, branch-and-cut algorithm) to solve different large MOPs.

One of the goals of the DOLPHIN project is to integrate the developed parallel models into software frameworks. Several frameworks for parallel distributed metaheuristics have been proposed in the literature. Most of them focus only either on evolutionary algorithms or on local search methods. Only few frameworks are dedicated to the design of both families of methods. On the other hand, existing optimization frameworks either do not provide parallelism at all or just supply at most one parallel model. In this project, a new framework for parallel hybrid metaheuristics is proposed, named *Parallel and Distributed Evolving Objects (ParadisEO)* based on EO. The framework provides in a transparent way the hybridization mechanisms presented in the previous section, and the parallel models described in the next section. Concerning the developed parallel exact methods for MOPs, we will integrate them into well-known frameworks such as COIN.

3.2.1. Parallel models

According to the family of addressed metaheuristics, we may distinguish two categories of parallel models: parallel models that manage a single solution, and parallel models that handle a population of solutions. The major single solution-based parallel models are the following: the *parallel neighborhood exploration model* and the *multi-start model*.

- *The parallel neighborhood exploration model* is basically a "low level" model that splits the neighborhood into partitions that are explored and evaluated in parallel. This model is particularly interesting when the evaluation of each solution is costly and/or when the size of the neighborhood is large. It has been successfully applied to the mobile network design problem (see Application section).
- *The multi-start model* consists in executing in parallel several local searches (that may be heterogeneous), without any information exchange. This model raises particularly the following question: is it equivalent to execute k local searches during a time t than executing a single local search during $k \times t$? To answer this question we tested a multi-start Tabu search on the quadratic assignment problem. The experiments have shown that the answer is often landscape-dependent. For example, the multi-start model may be well-suited for landscapes with multiple basins.

Parallel models that handle a population of solutions are mainly: the *island model*, the *central model* and *the distributed evaluation of a single solution*. Let us notice that the last model may also be used with single-solution metaheuristics.

- In *the island model*, the population is split into several sub-populations distributed among different processors. Each processor is responsible of the evolution of one sub-population. It executes all the steps of the metaheuristic from the selection to the replacement. After a given number of generations (synchronous communication), or when a convergence threshold is reached (asynchronous communication), the migration process is activated. Then, exchanges of solutions between sub-populations are realized, and received solutions are integrated into the local sub-population.
- *The central (Master/Worker) model* allows us to keep the sequentiality of the original algorithm. The master centralizes the population and manages the selection and the replacement steps. It sends sub-populations to the workers that execute the recombination and evaluation steps. The latter returns back newly evaluated solutions to the master. This approach is efficient when the generation and evaluation of new solutions is costly.
- *The distributed evaluation model* consists in a parallel evaluation of each solution. This model has to be used when, for example, the evaluation of a solution requires access to very large databases (data mining applications) that may be distributed over several processors. It may also be useful in a multi-objective context, where several objectives have to be computed simultaneously for a single solution.

As these models have now been identified, our objective is to study them in the multi-objective context in order to use them advisedly. Moreover, these models may be merged to combine different levels of parallelism and to obtain more efficient methods [52], [55].

3.2.2. Goals

Our objectives focus on these issues are the following:

- *Design of parallel models for metaheuristics and exact methods for MOPs*: We will develop parallel cooperative metaheuristics (evolutionary algorithms and local search algorithms such as the Tabu search) for solving different large MOPs. Moreover, we are designing a new exact method, named PPM (Parallel Partition Method), based on branch and bound and branch and cut algorithms. Finally, some parallel cooperation schemes between metaheuristics and exact algorithms have to be used to solve MOPs in an efficient manner.
- *Integration of the parallel models into software frameworks*: The parallel models for metaheuristics will be integrated in the ParadisEO software framework. The proposed multi-objective exact methods must be first integrated into standard frameworks for exact methods such as COIN and BOB++. A *coupling* with ParadisEO is then needed to provide hybridization between metaheuristics and exact methods.
- *Efficient deployment of the parallel models on different parallel and distributed architectures including GRIDs*: The designed algorithms and frameworks will be efficiently deployed on non-dedicated networks of workstations, dedicated cluster of workstations and SMP (Symmetric Multi-processors) machines. For GRID computing platforms, peer to peer (P2P) middlewares (XtremWeb-Condor) will be used to implement our frameworks. For this purpose, the different optimization algorithms may be re-visited for their efficient deployment.

4. Application Domains

4.1. Smart grids

With the smart grid revolution, house energy consumption will play a significant role in the energy system. Home users are indeed responsible for a significant portion of the world's energy needs portion, but are totally inelastic with respect to the market (i.e. the energy demand does not follow the price of the energy itself). Thus, the whole energy generation and distribution system performance can be improved by optimizing the house energy management. Those problems are concerned by multiple objectives such as cost and users' comfort, and multiple decision makers such as end-users and energy operators. We propose a home automation system that can monitor appliance scheduling in order to simultaneously optimize the total energy cost and the customer satisfaction.

The key challenge is to propose new optimization models and new hybrid optimization algorithms to the demand side management of smart grids in a context of uncertainty and in the presence of several conflicting objectives. Those complex optimization problems are also characterized by the presence of both continuous and discrete variables.

4.2. Transportation and logistics

- **Scheduling problems under uncertainty:** The flow-shop scheduling problem is one of the most well-known problems from scheduling. However, most of the works in the literature use a deterministic single-objective formulation. In general, the minimized objective is the total completion time (makespan). Many other criteria may be used to schedule tasks on different machines: maximum tardiness, total tardiness, mean job flowtime, number of delayed jobs, maximum job flowtime, etc. In the DOLPHIN project, a bi-criteria model, which consists in minimizing the makespan and the total tardiness, is studied. A bi-objective flow-shop problem with uncertainty on the duration, minimizing in addition the maximum tardiness, is also studied. It allows us to develop and test multi-objective (and not only bi-objective) optimization methods under uncertanty.
- **Routing problems under uncertainty:** The vehicle routing problem (VRP) is a well-known problem and it has been studied since the end of the fifties. It has a lot of practical applications in many industrial areas (ex. transportation, logistics, etc). Existing studies of the VRP are almost all concerned with the minimization of the total distance only. The model studied in the DOLPHIN project introduces a second objective, whose purpose is to balance the length of the tours. This new criterion is expressed as the minimization of the difference between the length of the longest tour and the length of the shortest tour. Uncertainty on the demands has also been introduced in the model.

4.3. Bioinformatics and Health care

Bioinformatic research is a great challenge for our society and numerous research entities of different specialities (biology, medical or information technology) are collaborating on specific themes.

4.3.1. Genomic and post-genomic studies

Previous studies of the DOLPHIN project mainly deal with genomic and postgenomic applications. These have been realized in collaboration with academic and industrial partners (IBL: Biology Institute of Lille; IPL: Pasteur Institute of Lille; IT-Omics firm).

First, genomic studies aim at analyzing genetic factors which may explain multi-factorial diseases such as diabetes, obesity or cardiovascular diseases. The scientific goal was to formulate hypotheses describing associations that may have any influence on diseases under study.

Secondly, in the context of post-genomic, a very large amount of data are obtained thanks to advanced technologies and have to be analyzed. Hence, one of the goals of the project was to develop analysis methods in order to discover knowledge in data coming from biological experiments.

These problems can be modeled as classical data mining tasks (Association rules, feature selection). As the combinatoric of such problems is very high and the quality criteria not unique, we proposed to model these problems as multi-objective combinatorial optimization problems. Evolutionary approaches have been adopted in order to cope with large scale problems.

Nowadays the technology is still going fast and the amount of data increases rapidly. Within the collaboration with Genes Diffusion, specialized in genetics and animal reproduction for bovine, swine, equine and rabbit species, we study combinations of Single Nucleotide Polymorphisms (SNP) that can explain some phenotypic characteristics. Therefore feature selection for regression is addressed using metaheuristics.

4.3.2. Optimization for health care

The collaboration with the Alicante company, a major actor in the hospital decision making, deals with knowledge extraction by optimization methods for improving the process of inclusion in clinical trials. Indeed, conducting a clinical trial, allowing for example to measure the effectiveness of a treatment, involves selecting a set of patients likely to participate to this test. Currently existing selection processes are far from optimal, and many potential patients are not considered. The objective of this collaboration consists in helping the practitioner to quickly determine if a patient is interesting for a clinical trial or not. Exploring different data sources (from a hospital information system, patient data...), a set of decision rules have to be generated. For this, approaches from multi-objective combinatorial optimization are implemented, requiring extensive work to model the problem, to define criteria optimization and to design specific optimization methods.

4.3.3. Molecular sampling and docking on large hybrid clusters

A Phd thesis is started in September 2015 in this context in collaboration with UMONS and University of Strasbourg. Flexible molecular docking is a very complex combinatorial opitmization problem especially when two components (ligand and protein) involved in the mechanism are together flexible. To deal in a reasonable time with such highly combinatorial process approximate optimization methods and massively parallel computing are absolutely The focus of the Ph.D thesis is on the flexibility-aware modeling and the design and implementation of near-approached optimization methods for solving the docking problem on large hybrid clusters including GPU accelerators and MIC coprocessors.

5. Highlights of the Year

5.1. Highlights of the Year

- Patent with the company Beckman: the invention relates to the handling of samples of biological material. In one aspect, the invention relates to optimization techniques for aliquoting such biological samples in a manner which accounts for various conditions and requirements as they may exist when the samples are to be processed.

6. New Software and Platforms

6.1. COCO

COmparing Continuous Optimizers
KEYWORDS: Benchmarking - Numerical optimization - Black-box optimization - Stochastic optimization
SCIENTIFIC DESCRIPTION

COmparing Continuous Optimisers (COCO) is a tool for benchmarking algorithms for black-box optimisation. COCO facilitates systematic experimentation in the field of continuous optimization. COCO provides: (1) an experimental framework for testing the algorithms, (2) post-processing facilities for generating publication quality figures and tables, (3) LaTeX templates of articles which present the figures and tables in a single document.

The COCO software is composed of two parts: (i) an interface available in different programming languages (C/C++, Java, Matlab/Octave, R, Python) which allows to run and log experiments on multiple test functions testbeds of functions (noisy and noiseless) are provided (ii) a Python tool for generating figures and tables that can be used in the LaTeX templates.

FUNCTIONAL DESCRIPTION

The Coco Platform provides the functionality to automatically benchmark optimization algorithms for unbounded, unconstrained optimization problems in continuous domains. Benchmarking is a vital part of algorithm engineering and a necessary path to recommend algorithms for practical applications. The Coco platform releases algorithm developers and practitioners alike from (re-)writing test functions, logging, and plotting facilities by providing an easy-to-handle interface in several programming languages. The Coco platform has been developed since 2007 and has been used extensively within the "Blackbox Optimization Benchmarking (BBOB)" workshop series since 2009. Overall, 140+ algorithms and algorithm variants by contributors from all over the world have been benchmarked with the platform so far and all data is publicly available for the research community). A new extension towards bi-objective problems will be used for the BBOB-2016 workshop at GECCO.

- Participants: Dimo Brockhoff, Arnaud Liefooghe, Thanh-Do Tran, Nikolaus Hansen, Anne Auger, Marc Schoenauer, Ouassim Ait Elhara, Asma Atamna, Tea Tusar and Dejan Tusar
- Partners: Université technique de Dortmund - Université technique de Prague
- Contact: Dimo Brockhoff
- URL: https://github.com/numbbo/coco

6.2. ParadisEO

KEYWORD: Metaheuristics, multi-objective optimization, Parallel metaheuristics

SCIENTIFIC DESCRIPTION

ParadisEO (PARallel and DIStributed Evolving Objects) is a C++ white-box object-oriented framework dedicated to the flexible design of metaheuristics. Based on EO, a template-based ANSI-C++ compliant evolutionary computation library, it is composed of four modules: * Paradiseo-EO provides tools for the development of population-based metaheuristic (Genetic algorithm, Genetic programming, Particle Swarm Optimization (PSO)...) * Paradiseo-MO provides tools for the development of single solution-based metaheuristics (Hill-Climbing, Tabu Search, Simulated annealing, Iterative Local Search (ILS), Incremental evaluation, partial neighborhood...) * Paradiseo-MOEO provides tools for the design of Multi-objective metaheuristics (MO fitness assignment shemes, MO diversity assignment shemes, Elitism, Performance metrics, Easy-to-use standard evolutionary algorithms...) * Paradiseo-PEO provides tools for the design of parallel and distributed metaheuristics (Parallel evaluation, Parallel evaluation function, Island model) Furthermore, ParadisEO also introduces tools for the design of distributed, hybrid and cooperative models: * High level hybrid metaheuristics: coevolutionary and relay model * Low level hybrid metaheuristics: coevolutionary and relay model

FUNCTIONAL DESCRIPTION

Paradiseo is a software framework for metaheuristics (optimisation algorithms aimed at solving difficult optimisation problems). It facilitates the use, development and comparison of classic, multi-objective, parallel or hybrid metaheuristics.

- Partners: Université Lille 1
- Contact: El-Ghazali Talbi
- URL: http://paradiseo.gforge.inria.fr/

6.3. VRPsolve

KEYWORDS: C++ - Mobile Computing, Transportation - Optimization

- Participants: Clive Ferret-Canape, Arnaud Liefooghe and Sebastien Verel
- URL: http://gforge.inria.fr/projects/vrpsolve

6.4. Platform Grid'5000

The Grid'5000 experimental platform is a scientific instrument to support computer science research related to distributed systems, including parallel processing, high performance computing, cloud computing, operating systems, peer-to-peer systems and networks. It is distributed on 10 sites in France and Luxembourg, including Lyon. Grid'5000 is a unique platform as it offers to researchers many and varied hardware resources and a complete software stack to conduct complex experiments, ensure reproducibility and ease understanding of results.

- Participants: F. Desprez, F. Huet, E. Jeannot, Y. Jegou, A. Lebre, L. Lefevre, F. Loui, D. Margery, N. Melab, J-M. Menaud, P. Neyron, L. Nussbaum, C. Perez, J-M. Pierson, O. Richard., S. Varette.
- Contact: Frédéric Desprez
- URL: https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home

7. New Results

7.1. Optimization under uncertainty

Participants: El-Ghazali Talbi, Raca Todosijevic, Oumayma Bahri (externel collaborators: Nahla BenAmor - Univ. Tunis, Tunisia, J. Puente, C. R. Vela, I. Gonzalez-Rodriguez - Univ. Oviedo Spain)

At the problem level, the sources of uncertainty are due to many factors such as the environment parameters of the model, the decision variables and the objective functions. Examples of such uncertainties can be the demand and travel times in vehicle routing problems, the execution time in scheduling problems, the wind or solar production in energy power systems, the price of resources in manufacturing, and the mechanical properties of a structure. Then, we need precise and efficient modeling and resolution approaches which are robust and non-sensitive to those uncertainties. The appeal of optimization under uncertainty is that its performance results remain relatively unchanged when exposed to uncertain data.

We have considered the fuzzy job shop, a job shop scheduling problem with uncertain processing times modelled as triangular fuzzy numbers. While the usual approaches to solving this problem involve adapting existing metaheuristics to the fuzzy setting, we have proposed instead to follow the framework of simheuristics from stochastic optimisation. More precisely, we integrate the simulation of possible realisations of the fuzzy problem with a genetic algorithm that solves the deterministic job shop. We test the resulting method, simGA, on a testbed of 23 benchmark instances and obtain results that suggest that this is a promising approach to solving problems with uncertainty by means of metaheuristics [38].

7.2. Indicator-based Multiobjective Optimization

Participants: Bilel Derbel, Arnaud Liefooghe (external collaborators: Matthieu Basseur, Adrien Goëffon, Univ. Angers, France)

A large spectrum of quality indicators has been proposed so far to assess the performance of discrete Pareto set approximations in multiobjective optimization. Such indicators assign, to any solution set, a real-value reflecting a given aspect of approximation quality. This is an important issue in multiobjective optimization, not only to compare the performance and assets of different approximate algorithms, but also to improve their internal selection mechanisms. In [37], we adopt a statistical analysis to experimentally investigate by how much a selection of state-of-the-art quality indicators agree with each other for a wide range of Pareto set approximations from well-known two- and three-objective continuous benchmark functions. More particularly, we measure the correlation between the ranking of low-, medium-, and high-quality limited-size approximation sets with respect to inverted generational distance, additive epsilon, multiplicative epsilon, R2, R3, as well as hypervolume indicator values. Since no pair of indicators obtains the same ranking of approximation sets, we confirm that they emphasize different facets of approximation quality. More importantly, our statistical analysis allows the degree of compliance between these indicators to be quantified.

Subset selection constitutes an important stage of any evolutionary multiobjective optimization algorithm when truncating the current approximation set for the next iteration. This appears to be particularly challenging when the number of solutions to be removed is large, and when the approximation set contains many mutually non-dominating solutions. In particular, indicator-based strategies have been intensively used in recent years for that purpose. However, most solutions for the indicator-based subset selection problem are based on a very simple greedy backward elimination strategy. We experiment additional heuristics that include a greedy forward selection and a greedy sequential insertion policies, a first-improvement hill-climbing local search, as well as combinations of those. We evaluate the effectiveness and the efficiency of such heuristics in order to maximize the enclosed hypervolume indicator of candidate subsets during a hypothetical evolutionary process, or as a post-processing phase. Our experimental analysis, conducted on randomly generated as well as structured two-, three- and four-objective mutually non-dominated sets, allows us to appreciate the benefit of these approaches in terms of quality, and to highlight some practical limitations and open challenges in terms of computational resources.

7.3. Decomposition-based Multiobjective Optimization

Participants: Bilel Derbel, Arnaud Liefooghe (external collaborators: Hernan Aguirre and Kiyoshi Tanaka, Shinshu Univ., Japan; Qingfu Zhang, City Univ., Hong Kong)

It is generally believed that local search (LS) should be used as a basic tool in multi-objective evolutionary computation for combinatorial optimization. However, not much effort has been made to investigate how to efficiently use LS in multi-objective evolutionary computation algorithms. In [28], we study some issues in the use of cooperative scalarizing local search approaches for decomposition-based multiobjective combinatorial optimization. We propose and study multiple move strategies in the MOEA/D framework. By extensive experiments on a new set of bi-objective traveling salesman problems with tunable correlated objectives, we analyze these policies with different MOEA/D parameters. Our empirical study has shed some insights about the impact of the Ls move strategy on the anytime performance of the algorithm.

7.4. Learning and Adaptation for Landscape-aware Algorithm Design

Participants: Bilel Derbel, Arnaud Liefooghe (external collaborators: Hernan Aguirre, Fabio Daolio, Miyako Sagawa and Kiyoshi Tanaka, Shinshu Univ., Japan; Cyril Fonlupt, Christopher Jankee and Sébastien Verel, Univ. Littoral, France)

In [13], we attempt to understand and to contrast the impact of problem features on the performance of randomized search heuristics for black-box multi-objective combinatorial optimization problems. At first, we measure the performance of two conventional dominance-based approaches with unbounded archive on a benchmark of enumerable binary optimization problems with tunable ruggedness, objective space dimension, and objective correlation (ρMNK-landscapes). Precisely, we investigate the expected runtime required by a global evolutionary optimization algorithm with an ergodic variation operator (GSEMO) and by a neighborhood-based local search heuristic (PLS), to identify a $(1+\varepsilon)$-approximation of the Pareto set. Then, we define a number of problem features characterizing the fitness landscape, and we study their intercorrelation and their association with algorithm runtime on the benchmark instances. At last, with a mixed-effects multi-linear regression we assess the individual and joint effect of problem features on the performance of both algorithms, within and across the instance classes defined by benchmark parameters. Our analysis reveals further insights into the importance of ruggedness and multi-modality to characterize instance hardness for this family of multi-objective optimization problems and algorithms.

Designing portfolio adaptive selection strategies is a promising approach to gain in generality when tackling a given optimization problem. However, we still lack much understanding of what makes a strategy effective, even if different benchmarks have been already designed for these issues. In [35], we propose a new model based on fitness cloud allowing us to provide theoretical and empirical insights on when an on-line adaptive strategy can be beneficial to the search. In particular, we investigate the relative performance and behavior of two representative and commonly used selection strategies with respect to static (off-line) and purely random approaches, in a simple, yet sound realistic, setting of the proposed model.

In evolutionary multi-objective optimization, variation operators are crucially important to produce improving solutions, hence leading the search towards the most promising regions of the solution space. In [39], we propose to use a machine learning modeling technique, namely random forest, in order to estimate, at each iteration in the course of the search process, the importance of decision variables with respect to convergence to the Pareto front. Accordingly, we are able to propose an adaptive mechanism guiding the recombination step with the aim of stressing the convergence of the so-obtained offspring. By conducting an experimental analysis using some of the WFG and DTLZ benchmark test problems, we are able to elicit the behavior of the proposed approach, and to demonstrate the benefits of incorporating machine learning techniques in order to design new efficient adaptive variation mechanisms.

7.5. Feature Selection using Tabu Search with Learning Memory: Learning Tabu Search

Participants: C. Dhaenens, L. Jourdan, M-E. Kessaci

Feature selection in classification can be modeled as a combinatorial optimization problem. One of the main particularities of this problem is the large amount of time that may be needed to evaluate the quality of a subset of features. We propose to solve this problem with a tabu search algorithm integrating a learning mechanism. To do so, we adapt to the feature selection problem, a learning tabu search algorithm originally designed for a railway network problem in which the evaluation of a solution is time-consuming. Experiments conducted show the benefit of using a learning mechanism to solve hard instances of the literature [hal-01370396v1].

7.6. MO-ParamILS: A Multi-objective Automatic Algorithm Configuration Framework

Participants: C. Dhaenens, L. Jourdan, M-E. Kessaci

Automated algorithm configuration procedures play an increasingly important role in the development and application of algorithms for a wide range of computationally challenging problems. Until very recently, these configuration procedures were limited to optimising a single performance objective, such as the running time or solution quality achieved by the algorithm being configured. However, in many applications there is more than one performance objective of interest. This gives rise to the multi-objective automatic algorithm configuration problem, which involves finding a Pareto set of configurations of a given target algorithm that characterises trade-offs between multiple performance objectives. In this work, we introduced MO-ParamILS, a multiobjective extension of the state-of-the-art single-objective algorithm configuration framework ParamILS, and demonstrated that it produces good results on several challenging bi-objective algorithm configuration scenarios compared to a base-line obtained from using a state-of-the-art single-objective algorithm configurator. [hal-01370392].

7.7. Parallel optimization methods revisited for multi-core and many-core (co)processors

Participants: J. Gmys and N. Melab

This contribution is a joint work with M. Mezmaz, E. Alekseeva and D. Tuyttens from University of Mons (UMONS) and T. C. Pessoa and F. H. De Carvalho Junior from Universidade Federal Do Cearà (UFC), Brazil.

On the road to exascale, coprocessors are increasingly becoming key building blocks of High Performance Computing platforms. In addition to their energy efficiency, these many-core devices boost the performance of multi-core processors. During 2016, we first have revisited the design and implementation of parallel Branch-and-Bound (B&B) algorithms using the work stealing paradigm on GPU accelerators [16][40], multi-GPU systems [17], multi-core processors [15] and MIC (Xeon Phi) coprocessors [20]. The challenge is to take into account the high irregular nature of the B&B algorithm and the hardware characteristics of GPU, Xeon Phi and multi-core (co)processors. Several work stealing strategies have been investigated while addressing several issues: host-device data transfer, thread divergence and data placement on the hierarchy of memories of the GPU and vectorization on Xeon Phi. The proposed approaches have been extensively experimented considering permutation-based optimization problems (e.g. FSP). The results reported in the cited papers demonstrate the efficiency of the many-core approaches compared to their multi-core counterpart. An extension of the proposed approaches to large hybrid clusters, including multi-core and many-core (co)processors is already started in [27].

The second part of the contribution consists in proposing a new hyper-heuristic (generalized GRASP) together with its parallelization for multi-core processors [11]. A cost function based on a bounding operator (used in B&B) is integrated to GRASP for the first time. Multi-core computing is used to investigate 315 GRASP configurations. In order to improve the performance of the local search procedure used in GRASP, we have proposed in [33] an original vectorization of the cost function of the makespan of FSP on Xeon Phi coprocessors. The reported results show that speed-ups up to 4.5 can be achieved compared to a non-vectorized apprpoach.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Beckman (2015-2018): the goal of this contract concerns the strategic and operational planning for medical laboratories (Phd of Sohrab Faramarzi).
- Strat&Logic (2012-2016): the objective of this CIFRE contract is the optimization of economic decisions in a competitive business management simulator (Phd of S. Dufourny).
- PIXEO (2014-2018): the objective of this bilateral project is the predictive models and knowledge extraction for insurance web comparator (Phd of A-L. Bedenel).
- Alicante (2014-2017): the objective of this CIFRE contract is the design of new optimization methods to extract knowledge from hospital data (Phd of M. Vandromme)
- Intel (2015-2016) Bilateral academic and research partnership between Université Lille 1 and Intel. In this context, Intel provides Lille 1 with training and technical support for the dissemination of its activities related to High Performance Computing.
- Nvidia (2016) Nvidia GPU Research Center, (see: https://developer.nvidia.com/academia/centers/universit%C3%A9-lille-1).

8.2. Bilateral Grants with Industry

- Intel 2015-2016 Intel has supported with a budget equivalent to 22Keuros the acquisition of a cluster of 2 multi-core servers and 8 Intel Xeon Phi coprocessors. The objective is to develop research and teaching on multi and many-core computing on coprocessors. The hybrid cluster has been deployed in 2016.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- CPER “data" (2015-2020): co-leader of a workpackage “Research infrastructures". The objective is to support research related to data science including high performance computing for combinatorial optimization using the Grid’5000 grid infrastructure.
- ELSAT (2015-2019) of CPER (Contrat Plan Etat Région) : transversal research action “Planning and scheduling of maintenance logistics in transportation”.

9.2. National Initiatives

9.2.1. ANR

- ANR project Modèles Numériques “NumBBO - Analysis, Improvement and Evaluation of Numerical Blackbox Optimizers” (2012-2016) in collaboration with Inria Saclay, TAO team, Ecole des Mines de St. Etienne, CROCUS team, and TU Dortmund University, Germany (2012-2016).
- ANR project TECSAN (Technologies pour la Santé) “ClinMine - Optimisation de la prise en Charge des Patients à l’Hôpital”, in collaboration with University Lille 1, Université Lille 2, Inria, CHRU Lille, CHICL, Alicante (7 partners) (2014-2017) - Coordinator -
- Bilateral ANR/RGC France/Hong Kong PRCI “Big Multiobjective Optimization” (2016-2021) in collaboration with City University of Hong Kong.
- PGMO project “Towards a Complexity Theory for Black-Box Optimization”, together with Carola Doerr (CNRS, LIP6), Benjamin Doerr (Ecole Polytechnique), Anne Auger, Nikolaus Hansen (both Inria Saclay), Timo Koetzing (University of Jena, Germany), Johannes Lengler (ETH Zurich, Switzerland), and Jonathan Rowe (The University of Birmingham, UK), (2014-2016)
- PGMO project “Demand side management in smart grids”, together with EDF, (2015-2017).

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

Program: H2020

Project acronym: SYNERGY

Project title: Synergy for Smart Multi-Objective Optimisation

Duration: 02 2016 - 01 2019

Coordinator: Jožef Stefan Institute (JSI), Ljubljana, Slovenia

Other partners: University of Lille (France), Cologne University of Applied Sciences (Germany)

Abstract: Many real-world application areas, such as advanced manufacturing, involve optimisation of several, often time-consuming and conflicting objectives. For example, they require the maximisation of the product quality while minimising the production cost, and rely on demanding numerical simulations in order to assess the objectives. These, so-called multi-objective optimisation problems can be solved more efficiently if parallelisation is used to execute the simulations simultaneously and if the simulations are partly replaced by accurate surrogate models.

9.3.2. Collaborations with Major European Organizations

University of Luxembourg: (Luxembourg)

Energy aware scheduling in Cloud computing systems

University of Oviedo: (Spain)

Optimization under uncertainty for fuzzy flow shop scheduling

University of Elche and University of Murcia: (Spain)

Matheuristics for DEA

9.4. International Initiatives

9.4.1. Inria International Labs

- LIRIMA Afrique: Equipe associér avec l'EMI (Ecole Mohammadia d'Ingénieurs), Morocco

9.4.2. Inria Associate Teams Not Involved in an Inria International Labs

9.4.2.1. MOHA

Title: Mixed Multi-objective Optimization using Hybrid Algorithms: Application to smart grids

International Partner (Institution - Laboratory - Researcher):

Ecole Mohammadia d'Ingénieurs (Morocco) - LERMA (Laboratoire d'Etudes et de Recherches en Mathématiques Appliquées) - Rachid Ellaia

Start year: 2016

See also: https://ocm.univ-lille1.fr/ talbi/momh

The key challenge of this project is to propose new optimization models and new hybrid algorithms to the demand side management of smart grids in a context of uncertainty and in the presence of several conflicting objectives.

Those complex optimization problems are also characterized by the presence of both continuous and discrete variables. We need to design new efficient optimization algorithms combining state of the art exact and metaheuristic algorithms from the global optimization and combinatorial optimization communities

9.4.2.2. s3-bbo

Title: Threefold Scalability in Any-objective Black-Box Optimization (s3-bbo)

International Partner (Institution - Laboratory - Researcher):

Shinshu University, Japan

Duration: 2015-2017

See also: http://francejapan.gforge.inria.fr/doku.php?id=associateteam

The main scientific goals of this collaboration is to theoretically derive, analyze, design, and develop scalable evolutionary and other stochastic local search algorithms for large-scale optimization considering three different axes of scalability: (i) decision space, (ii) objective space, and (iii) availability of distributed and parallel computing resources. This research will allow us to design, control, predict, analyze and optimize parameters of recent complex, large-scale, and computationally expensive systems, providing the basic support for problem solution and decision-making in a variety of real world applications. For single-objective continuous optimization, we want to theoretically derive variants of the state-of-the-art CMA-ES with linear time and space complexity scalings with respect to the number of variables. We will exploit the information geometry framework to derive updates using parametrization of the underlying family of probability distribution involving a linear number of components. The challenges are related to finding good representations that are theoretically tractable and meaningful. For the design of robust algorithms, implementing the derived updates, we plan to follow the same approach as for the design of CMA-ES. For multi- and many-objective optimization, we will start by characterizing and defining new metrics and methodologies to analyze scalability in the objective space and in terms of computational resources. The first challenge is to accurately measure the impact of adding objectives on the search behavior and on the performance of evolutionary multi- and many- objective optimization (EMyO) algorithms. The second challenge is to investigate the new opportunities offered by large-scale computing platforms to design new effective algorithms for EMyO optimization. To this end, we plan to follow a feature-based performance analysis of EMyO algorithms, to design new algorithms using decomposition-based approaches, and to investigate their mapping to a practical parallel and distributed setting.

9.4.3. Inria International Partners

9.4.3.1. Declared Inria International Partners

- Memorandum of Understanding between Shinshu University (Japan) and Inria, signed on March 2014

9.4.3.2. Informal International Partners

- University of Coimbra, Portugal
- University of Manchester, U.K.
- Collaboration with Université de Mons (UMONS). The collaboration consists mainly in the joint supervision of the Ph.D thesis of Jan Gmys started in 2014.

9.4.4. Participation in Other International Programs

- JSPS-MEXT project on Evolutionary multi-objective optimization, landscape analysis, and search performance, with Shinshu University, Japan (2013—2016)

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Hernan Aguirre, Shinshu University, Japan
- Fabio Daolio, University of Stirling, U.K.
- Luis Paquete, University of Coimbra, Portugal
- Kiyoshi Tanaka, Shinshu University, Japan
- Saúl Zapotecas-Martínez, Shinshu University, Japan
- Qingfu Zhang, City University, Hong Kong
- Dr. Myriam Delgado (Federal University of Technology of Paraná, Brazil), 1 week, April 2016
- Prof. Fred Glover (University of Colorado, USA), 1 month, Nov 2016
- Dr Lakhdar Loukil from Université d'Oran, Algeria (January 18-22, 2016).

9.5.1.1. Internships

- Oliver Cuate, CINVESTAV, Mexico
- Miyako Sagawa, Shinshu University, Japan

9.5.2. Visits to International Teams

9.5.2.1. Sabbatical programme

- E-G. Talbi has a one-year sabbatical program for 2016 and 2017.

9.5.2.2. Research Stays Abroad

- E-G. Talbi: University of Florida, USA, 1 month, 2016.
- E-G. Talbi: University of Colorado, USA, 1 month, 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- N. Melab: Chair of the HPCS'2016 workshop (Parallel Optimization using / for Multi and Many-core High Performance Computing) organized in conjunction with HPCS'2016, Innshbruck, Austria, June 7th 2015.
- N. Melab: Chair of 5 simulation and HPC-related seminars at Lille 1 oct-dec. 2016 (CENAERO, Intel, Atos-Bull, FFT, UPMC).
- E-G. Talbi: General chair of META'2016 Int. Conf. on Metaheuristics and Nature Inspired Computing, Marrakech, Morocco, Oct 2016, 105 participants.
- E-G. Talbi, Program co-chair of HM'2016 Int. Conf on Hybrid Metaheuristics, Exeter, UK, May 2016.

10.1.1.2. Member of the Organizing Committees

- D. Brockhoff: co-organizer of the Surrogate-Assisted Multi-Criteria Optimization workshop at the Lorentz Center in Leiden, The Netherlands, Feb/Mar 2016
- D. Brockhoff: co-organizer of the Blackbox Optimization Benchmarking workshop (BBOB-2016) at GECCO in Denver, CO, USA
- CEC 2016 special session entitled "Advances in Decomposition-based Evolutionary Multiobjective Optimization", Vancouver, Canada, organized by Saul Zapotecas Martinez, Bilel Derbel, Qingfu Zhang, Carlos A. Coello Coello, July 2016
- E-G. Talbi: organisation of META'2016 Int. Conf. on Metaheuristics, Marrakech, Morocco, Oct 2016.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- E-G. Talbi, HM'2016
- E-G. Talbi, META'2016

10.1.2.2. Member of the Conference Program Committees

- CEC -IEEE Congress on Evolutionary Computation 2016
- CIBCB - IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology 2016
- GECCO conference 2016
- HM 2016
- ICORES 2016
- LION Conference 2016
- MICAI 2016
- MIM 2016
- MOD 2016
- PPSN 2016
- ROADEF 2016
- GECCO conference 2016
- IEEE Congress on Evolutionary Computation (CEC), Vancouver, Canada, July 24-29, 2016
- The ACM Genetic and Evolutionary Computation Conference (GECCO), Denver, Colorado, USA, July 20-24, 2016
- IEEE International Workshop on Nature Inspired Distributed Computing (IPDPS/NIDISC'2016), Chicago, Illinois, USA, May 23-27, 2016
- IEEE Intl. Workshop on Parallel Computing and Optimization (IPDPS/PCO), Chicago, Illinois, USA, May 23-27, 2016

- Grid'5000 winter school, Grenoble, France, February 2-5, 2016
- Colloque sur l'Optimisation et les Systèmes d'information (COSI), Sétif, Algérie, May 30 - June 1, 2016
- Intl. Conf. on Contemporary Computing (IC3), Noida, India, Aug. 11-13, 2016
- The 2nd Intl. Conf. on Cloud Computing Technologies and Applications (CloudTech), Marrakesh, Morocco, May 24-26, 2016.
- 8th IEEE Intl. Conf. on Cloud Computing Technology and Science (CloudCom), Luxembourg, Dec. 12-15, 2016
- PPSN 2016: 14th International Conference on Parallel Problem Solving from Nature (Edinburgh, UK, 2016)
- GECCO 2016: Genetic and Evolutionary Computation Conference, Evolutionary Combinatorial Optimization and Metaheuristics (ECOM) track (Denver, USA, 2016)
- CEC 2016: IEEE Congress on Evolutionary Computation (Vancouver, Canada, 2016)
- EvoCOP 2016: 16th European Conference on Evolutionary Computation in Combinatorial Optimisation (Porto, Portugal, 2016)

10.1.2.3. Reviewer

- Dimo Brockhoff: CEC'2016, GECCO'2016 (EMO track), PPSN'2016, FOGA'2017, EMO'2017

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- L. Jourdan: Review Editor Frontiers in Big Data
- N. Melab: Guest Editor (in collaboration with M. Mezmaz) of a special on Multi/Many-core computing for parallel Metaheuristics in Wiley Concurrency and Computation: Practice and Experience, April 2016.
- N. Melab: Guest Editor (in collaboration with A. Zomaya and I. Chakroun) of a special on Parallel Optimization using/for Multi and Many-core High Performance Computing in Journal of Parallel and Distributed Computing (JPDC), 2016.
- E-G. Talbi : Editor of the Journal « Computers and Industrial Engineering (CAIE, Elsevier)» Area «Computational Intelligence».

10.1.3.2. Reviewer - Reviewing Activities

- IEEE Transactions on Evolutionary Computation, Evolutionary Computation, Journal of Heuristics, Artificial Intelligence Journal
- Applied Soft Computing
- Computers in Biology and Medecine
- Computers & Industrial Engineering
- Computers & Operations Research
- EJOR European Journal of Operational Research
- IEEE Transaction on Evolutionary Computation
- International Journal of Metaheuristics
- International Journal of Molecular Sciences
- International Journal of Production research
- International Transactions in Operational
- JOH Journal of Heuristics
- JOCO Journal of Combinatorial Optimization
- JPDC Journal of Parallel and Distributed Computing

- Nature Scientific Report
- Soft Computing (SOCO)
- Transactions on Computational Biology and Bioinformatics
- ACM Computing Surveys
- Computation and Concurrency: Practice and Experience (CCPE)
- Parallel Processing Letters
- Parallel Computing
- Journal of Parallel and Distributed Computing (JPDC)
- 4OR: A Quarterly Journal of Operations Research (Springer)
- ASOC: Applied Soft Computing (Elsevier)
- CAIE: Computers & Industrial Engineering (Elsevier)
- ITOR: International Transactions in Operational Research (Wiley)
- NEUCOM: Neurocomputing (Elsevier)

10.1.4. Invited Talks

- D. Brockhoff: invited talk on multiobjective optimization, MEXICO/Mascot-Num meeting, Nov 2016, Nantes
- D. Brockhoff: invited tutorial at GECCO'2016 on Evolutionary Multiobjective Optimization, Jul 2016, Denver, CO, USA
- B.Derbel and A. Liefooghe: Designing and understanding EMO algorithms, Invited talk, City University, Hong Kong, November 2016
- A. Liefooghe: Fitness landscape analysis, problem features and performance prediction for multi-objective optimization, Workshop on Landscape-aware heuristic search (PPSN 2016), Edinburgh, UK, September 2016 (joint work with Fabio Daolio, Sébastien Verel, Hernan Aguirre, and Kiyoshi Tanaka)
- L. Jourdan, "Combinatorial optimization for Bioinformatics", invited talk (1day), summer school of Bioinformatics, Angers, 2016
- L. Jourdan, "The emerging use of Optimization methods for Datamining in Big Data", invited talk, summer school of Cyber-Physical Systems (CPS), Toulouse, 2016.
- L. Jourdan, "Modélisation et optimisation multi-objectif pour l'extraction de connaissances Le cas des applications médicales", Mars 2016, Séminaire Expert, Worldline.
- C. Dhaenens "Exemple de collaboration réussie entre l'entreprise et le monde de la recherche", CCI Grand Lille, Fev. 2016.
- N. Melab: Tutorial on Grid'5000, Arcus international project "E2D2", May 2016, Université Lille 1.
- E-G. Talbi: Multi-objective metaheuristics, Invited seminar, Colorado State University, Fort Colins, Colorado, USA, Mar 2016.
- E-G. Talbi: Optimization under uncertainty, Invited seminar, Univeridad Elche, Elche, Spain, Apr 2016.
- E-G. Talbi: Parallel evolutionary algorithms for multi-objective optimization, Keynote speaker BIOMA'2016 7th Int. Conf. on Bioinspired Optimization Methods and their Applications, Bled, Slovenia, May 2016.
- E-G. Talbi: Parallel metaheuristics, Invited seminar, CINVESTAV, Mexico, Sept 2016.
- E-G. Talbi: Combining metaheuristics with mathematical programming and data mining, Keynote speaker, NEO'2016 Int. Workshop on Numerical and Evolutionary Optimization, Tlalnepantla, Mexico, Sept 2016.

- E-G. Talbi: A survey of hybrid metaheuristics with exact methods and machine learning, Tutorial, META'2016 Int. Conf. on Metaheuristics and Nature Inspired computing, Marrakech, Morocco, Oct 2016.

10.1.5. Leadership within the Scientific Community

- L. Jourdan : Co-president of the working group “ATOM: Multi-objective optimization”, GDR RO.
- L. Jourdan, A. Liefooghe : Secretary of the association “Artificial Evolution” (EA).
- C. Dhaenens: member of the scientific council of GDR RO (Operations research)
- C. Dhaenens: nominated member at Co-NRS, section 6 (National committee of CNRS)
- N. Melab: scientific leader of Grid'5000 (https://www.grid5000.fr) at Lille, Since 2004
- N. Melab: Chargé de Mission of High Performance Computing and Simulation at Université Lille 1, Since 2010
- E-G. Talbi : Co-president of the working group “META: Metaheuristics - Theory and applications”, GDR RO and GDR MACS.
- E-G. Talbi : Co-Chair of the IEEE Task force on Cloud Computing within the IEEE Computational Intelligence Society.

10.1.6. Scientific Expertise

- D. Brockhoff: external reviewer of a research proposal for the National Science Centre Poland
- N. Melab: Member of the advisory committee for the IT and management engineer training at Faculté Polytechnique de Mons
- E-G. Talbi : Expert for Qatar Foundation QNRF projects, 2016.

10.1.7. Research Administration

- C. Dhaenens: Vice-head of CRIStAL laboratory (Centre de Recherche en Informatique, Signal et Automatique de Lille), common to CNRS, University of Lille and Ecole Centrale de Lille, 430 people.
- L. Jourdan: member of the Bureau du Département de domaine Informatique pour l'école doctorale SPI, University of Lille
- N. Melab: Member of the steering committee of “Maison de la Simulation" at Université Lille 1
- E-G. Talbi, Coordinator of the International Relationships of Inria Lille Nord Europe.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master : Dimo Brockhoff, Introduction to Optimization, 18h ETD, M2 Apprentissage, Information et Contenu, U. Paris-Saclay, France
- Master : Dimo Brockhoff, Advanced Optimization, 18h ETD, M2 Apprentissage, Information et Contenu, U. Paris-Saclay, France
- Master : Dimo Brockhoff, Introduction to Optimization, 54h ETD, MSc in Data Sciences & Business Analytics, CentraleSupelec/ESSEC, France
- Master : Laetitia Jourdan, Business Intelligence, 30h, M1, University of Lille 1, France
- Master : Laetitia Jourdan, Datamining, 60h , M1, University of Lille 1, France
- Master : Laetitia Jourdan, Datawarehouse, 30h, M1, University of Lille 1, France
- Licence: Laetitia Jourdan : Informatique, 48h, L1 University of Lille 1, France
- Master: Laetitia Jourdan : Responsible of Master MIAGE Formation en Alternance, ,University of Lille 1, France

- Licence: Laetitia Jourdan: Co-responsible of Licence 1 Computer Science, University of Lille 1, France
- Engineering school : Clarisse Dhaenens, Graphs and Combinatorics, 80 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Clarisse Dhaenens, Operations Research, 70 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Clarisse Dhaenens, Algorithmics and programming, 45 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Clarisse Dhaenens, responsible of the 5th year of statistics and computer science department.
- Engineering school : Marie-Eléonore Kessaci, Graphs and Combinatorics, 44 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Marie-Eléonore Kessaci, Algorithmics and programming, 51 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Marie-Eléonore Kessaci, Databases, 71 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Marie-Eléonore Kessaci, Mathematics, 20 HeqTD, Polytech Lille, University Lille 1, France
- Engineering school : Marie-Eléonore Kessaci, responsible of the 3th year of statistics and computer science department.
- Master lecture: N. Melab, Supercomputing, 24h, Master 2, Université Lille 1, France
- Master lecture: N. Melab, Operations Research, 78h, Master 1, Université Lille 1, France
- Master leading: N. Melab, Co-head (with C. Chainais) of the master 2 of advanced scientific computing, U. Lille 1
- Licence: A. Liefooghe, Algorithmic and Data structure, 36h ETD, L2, Université de Lille 1, France
- Licence: A. Liefooghe, Algorithmic for Operations Research, 36h ETD, L3, Université de Lille 1, France
- Master: A. Liefooghe, Databases, 30h ETD, M1, Université de Lille 1, France
- Master: A. Liefooghe, Advanced Object-oriented Programming, 53h ETD, M2, Université de Lille 1, France
- Master: A. Liefooghe, Combinatorial Optimization, 10h ETD, M2, Université de Lille 1, France
- Master: A. Liefooghe, Multi-criteria Decision Aid and Optimization, 25h ETD, M2, Université de Lille 1, France
- A. Liefooghe is supervising the Master 2 MIAGE IPI-NT
- Master : Bilel Derbel, Combinatorial Optimization, 35h, M2, University Lille 1, France
- Master : Bilel Derbel, Grid Computing, 16h, M2, University Lille 1, France
- Master : Bilel Derbel, Parallel and Distributed Programming, 35h, M1, University Lille 1, France
- Master : Bilel Derbel, Algorithms and Applications, 28h, M1, University Lille 1, France
- Engineering school : El-Ghazali Talbi, Advanced optimization, 36h, Polytech'Lille, University Lille 1, France
- Engineering school : El-Ghazali Talbi, Data mining, 36h, Polytech'Lille, University Lille 1, France
- Engineering school : El-Ghazali Talbi, Operations research, 60h, Polytech'Lille, University Lille 1, France
- Engineering school : El-Ghazali Talbi, Graphs, 25h, Polytech'Lille, University Lille 1, France

10.2.2. Supervision

- PhD in progress: Gauvain Marquet, Mono-objective decomposiiton for multi-objective optimization, University Lille 1, Sep. 2014, Bilel Derbel and El-Ghazali Talbi
- PhD in progress: Maxence Vandromme, Datamining et optimisation combinatoire adaptés à la prévention et à l'orientation de patients, début : 1/06/2014, CIFRE with Alicante Co-direction : Clarisse Dhaenens and Laetitia Jourdan
- PhD in progress : Sylvain Dufourny, Optimisation de décisions économiques concurrentielles dans un simulateur de gestion d'entreprise, Novembre 2012, Clarisse Dhaenens
- PhD in progress : Aymeric Blot, Réagir et s'adapter à son environnement : Concevoir des méthodes autonomes pour l'optimisation combinatoire à plusieurs objectifs, september 2015, co-directed Laetitia Jourdan and Marie-Eléonore Marmion
- PhD in progress : Lucien Mousin, Exploiter la connaissance pour mieux optimiser, october 2015, co-directed Clarisse Dhaenens and Marie-Eléonore Marmion
- PhD in progress : AnneLise Bedenel, Classification supervisée et non supervisée en présence de descripteurs évoluant dans le temps. Application à la comparaison d'assurances en ligne, co-directed Laetitia Jourdan and Christophe Biernacki (Modal Inria Team)
- PhD (cotutelle in progress): Jan GMYS, Parallel Branch-and-Bound for solving permutation problems on multi- and many-core clusters, Nouredine Melab (Université Lille 1) and Daniel Tuyttens (UMONS, Belgium), Defense end of 2017
- PhD in progress : A. Q. Nguyen, Green scheduling on cloud computing systems, 11/2012, El-Ghazali Talbi and Pascal Bouvry
- PhD in progress : Oumayma Bahri, Fuzzy multi-objective optimization, 11/2013, El-Ghazali Talbi and Nahla Ben-Omar
- PhD in progress : Sohrab Faramarzi, Optimization of medical lab, 02/2016, El-Ghazali Talbi

10.2.3. Juries

- C. Dhaenens: PhD Thesis: B. Tounsi, “Contributions à la chaîne logistique e-commerce : Intégration dans l'e-fulfillment et tarification de services de livraison”, Université Lille 1, Dec. 2016.
- C. Dhaenens: HDR : L. Boudjeloud, “Approches coopératives et semi-interactives pour le traitement de données massives et temporelles”, Université de Lorraine, Dec. 2016.
- L. Jourdan: PhD Thesis: Métaheuristiques hybrides distribuées et massivement parallèles, de Omar ABDELKAFI Université de Haute Alsace, November 7th 2016 (Présidente de Jury)
- L. Jourdan: PhD Thesis: Le routage avec transbordement et collaboration, de Nicolas Danloup, Université de Béthune Artois, December 1st 2016 (Présidente de Jury) Hyperheuristics in Logistics, de Kassem DANACH, Ecole Centrale Lille, December 21st 2016 (Présidente de Jury).
- L. Jourdan: PhD Thesis: Contribution à la synthèse et l'optimisation multi-objectif par essaims particulaires de lois de commande robuste RST de système dynamique, de 'Riadh Madiouni', de l'Université Paris Est - Créteil, June 20th 2016 (Rapporteur)
- L. Jourdan: PhD Thesis: A dynamic programming operator for metaheuristics to solve vehicle routing problems with optional visits, de 'Leticia VARGAS' du LAAS-CNRS, June 24th 2016. (Rapporteur)
- L. Jourdan: PhD Thesis: Conception d'alliages par optimisation combinatoire multiobjectifs : thermodynamique prédictive, fouille de données, algorithmes génétiques et analyse décisionnelle de 'Edern Menou' Université de Nantes, October 19th 2016. (Rapporteur)
- HDR: Sebastien Verel, “Apport à l'analyse des paysages de fitness pour l'optimisation mono-objective et multi-objective", Université du Littoral - Côte d'Opale, December 12th, 2016.
- PhD thesis: Ania Kaci, “Conception d'une architecture extensible pour le calcul massivement parallèle", Université Paris-Est, December 14th, 2016.

- PhD thesis: Escobar Fernando, “High Performance Computing Archtectures based on Reconfigurable Platforms for Scientific Applications, Université de Mons, March 30th, 2016.
- PhD thesis: K. Lefrouni, “Contrôle de congestion dans les réseaux de communication”, EMI – Université Mohammed V Rabat, Maroc, Jan 2016.
- PhD thesis: S. Nielsen, “Diversity preserving genetic algorithms – Application to the inverted folding problem and analogous formulated benchmarks”, University of Luxembourg, Luxembourg, Feb 2016.
- PhD thesis: K. E. Vazquez Ortiz, “Advanced methods to solve the maximum parsimony problem”, Université d’Angers, France, June 2016.
- PhD thesis: Urrego Agudelo Lilliam, “A novel method for the approximation of risk of blackout in operational conditions”, Université Paris-Est, Créteil, Nov 2016.

10.3. Popularization

- Clarisse Dhaenens, Fanny Dufossé, Laetitia Jourdan, Marie-Eléonore Marmion: Operational research - for 2nde during integration week (June 2016)
- Laetitia Jourdan, Marie-Eléonore: Computer Unplugged, Numériqu’elle Day (November 2016)
- Laetitia Jourdan: Computer Unplugged, Primary School (December 2016)

11. Bibliography

Major publications by the team in recent years

[1] J.-C. BOISSON, L. JOURDAN, E.-G. TALBI.*Metaheuristics based de novo protein sequencing: A new approach*, in "Applied Soft Computing", 2011, vol. 11, n° 2, p. 2271-2278.

[2] C. DHAENENS, J. LEMESRE, E.-G. TALBI.*K-PPM: A new exact method to solve multi-objective combinatorial optimization problems*, in "European Journal of Operational Research", 2010, vol. 200, n° 1, p. 45-53.

[3] J. FIGUEIRA, A. LIEFOOGHE, E.-G. TALBI, A. P. WIERZBICKI.*A parallel multiple reference point approach for multi-objective optimization*, in "European Journal of Operational Research", 2010, vol. 205, n° 2, p. 390 - 400.

[4] N. JOZEFOWIEZ, F. SEMET, E.-G. TALBI.*Target Aiming Pareto Search and its application to the vehicle routing problem with route balancing*, in "Journal of Heuristics", 2007, vol. 13, p. 455-469.

[5] A. KHANAFER, F. CLAUTIAUX, E.-G. TALBI.*New lower bounds for bin packing problems with conflicts*, in "European Journal of Operational Research", 2010, vol. 2, n° 206.

[6] A. LIEFOOGHE, L. JOURDAN, E.-G. TALBI.*A software framework based on a conceptual unified model for evolutionary multiobjective optimization: ParadisEO-MOEO*, in "European Journal of Operational Research", 2010.

[7] A. LIEFOOGHE, L. PAQUETE, J. FIGUEIRA.*On local search for bi-objective knapsack problems*, in "Evolutionary Computation", 2013, vol. 21, n° 1, p. 179-196 [*DOI :* 10.1162/EVCO_A_00074], http://hal.inria.fr/hal-00676625.

[8] T. V. LUONG, N. MELAB, E.-G. TALBI.*GPU Computing for Parallel Local Search Metaheuristics*, in "IEEE Transactions on Computers", 2013, vol. 62, n^{o} 1, p. 173-185, http://hal.inria.fr/inria-00638805.

[9] M.-É. MARMION, L. JOURDAN, C. DHAENENS.*Fitness Landscape Analysis and Metaheuristics Efficiency*, in "Journal of Mathematical Modelling and Algorithms in Operations Research", 2013, vol. 12, n^{o} 1, p. 3-26 [*DOI :* 10.1007/S10852-012-9177-5], http://hal.inria.fr/hal-00807352.

[10] E.-G. TALBI.*Metaheuristics: From Design to Implementation*, Wiley, 2009.

Publications of the year

Articles in International Peer-Reviewed Journal

[11] E. ALEKSEEVA, M. MEZMAZ, D. TUYTTENS, N. MELAB.*Parallel multi-core hyper-heuristic GRASP to solve permutation flow-shop problem: Hyper-heuristique GRASP parallèle multi-coeur pour la résolution du flow-shop de permutation*, in "Concurrency and Computation: Practice and Experience", April 2016 [*DOI :* 10.1002/CPE.3835], https://hal.inria.fr/hal-01419060.

[12] D. C. CATTARUZZA, N. ABSI, D. FEILLET.*Vehicle routing problems with multiple trips*, in "4OR: A Quarterly Journal of Operations Research", 2016, https://hal-emse.ccsd.cnrs.fr/emse-01250603.

[13] F. DAOLIO, A. LIEFOOGHE, S. VEREL, H. AGUIRRE, K. TANAKA.*Problem Features vs. Algorithm Performance on Rugged Multi-objective Combinatorial Fitness Landscapes*, in "Evolutionary Computation", 2016 [*DOI :* 10.1162/EVCO_A_00193], https://hal.archives-ouvertes.fr/hal-01380612.

[14] F. DUFOSSÉ, B. UÇAR.*Notes on Birkhoff-von Neumann decomposition of doubly stochastic matrices*, in "Linear Algebra and its Applications", February 2016, vol. 497, p. 108–115 [*DOI :* 10.1016/J.LAA.2016.02.023], https://hal.inria.fr/hal-01270331.

[15] J. GMYS, R. LEROY, M. MEZMAZ, N. MELAB, D. TUYTTENS.*Work Stealing with Private Integer-Vector-Matrix Data Structure for Multi-core Branch-and-Bound Algorithms*, in "Concurrency and Computation: Practice and Experience", 2016 [*DOI :* 10.1002/CPE.3771], https://hal.inria.fr/hal-01248336.

[16] J. GMYS, M. MEZMAZ, N. MELAB, D. TUYTTENS.*A GPU-based Branch-and-Bound algorithm using Integer-Vector-Matrix data structure*, in "Parallel Computing", 2016 [*DOI :* 10.1016/J.PARCO.2016.01.008], https://hal.inria.fr/hal-01389471.

[17] J. GMYS, M. MEZMAZ, N. MELAB, D. TUYTTENS.*IVM-based parallel branch-and-bound using hierarchical work stealing on multi-GPU systems*, in "Concurrency and Computation: Practice and Experience", October 2016 [*DOI :* 10.1002/CPE.4019], https://hal.inria.fr/hal-01419072.

[18] M. GÉRARD, F. CLAUTIAUX, R. SADYKOV, M. GERARD.*Column generation based approaches for a tour scheduling problem with a multi-skill heterogeneous workforce*, in "European Journal of Operational Research", 2016 [*DOI :* 10.1016/J.EJOR.2016.01.036], https://hal.inria.fr/hal-01419205.

[19] S. JACQUIN, L. JOURDAN, E.-G. TALBI.*A multi-objective dynamic programming-based metaheuristic to solve a bi-objective unit commitment problem using a multi-objective decoder*, in "International Journal of Metaheuristics", 2016, vol. 5, n^{o} 1, p. 3-30 [*DOI :* 10.1504/IJMHEUR.2016.079104], https://hal.archives-ouvertes.fr/hal-01417960.

[20] N. MELAB, J. GMYS, M. MEZMAZ, D. TUYTTENS.*Multi-core versus Many-core Computing for Many-task Branch-and-Bound applied to Big Optimization Problems*, in "Future Generation Computer Systems", December 2016, 20, Accepted with minor revision, https://hal.inria.fr/hal-01419079.

[21] E.-G. TALBI.*Combining metaheuristics with mathematical programming, constraint programming and machine learning*, in "Annals of Operations Research", 2016, vol. 240, n^{o} 1, p. 171-215, https://hal.inria.fr/hal-01423492.

[22] T.-T. VU, B. DERBEL.*Parallel Branch-and-Bound in Multi-core Multi-CPU Multi-GPU Heterogeneous Environments*, in "Future Generation Computer Systems", March 2016, vol. 56, p. 95–109 [*DOI :* 10.1016/J.FUTURE.2015.10.009], https://hal.inria.fr/hal-01249124.

International Conferences with Proceedings

[23] O. BAHRI, E.-G. TALBI, N. BENAMOR.*Robust routes fort he fuzzy multi-objective vehicle routing problem*, in "MIM‘2016 8th IFAC Conference on Manufacturing Modelling Management and Control", Troyes, France, 2016, https://hal.inria.fr/hal-01423498.

[24] O. BAHRI, E.-G. TALBI, N. BENAMOR.*β-Robustness approach for fuzzy multi-objective problems*, in "IPMU‘2016 16th Int. Conf. On Information Processing and Management on Uncertainty in Knowledge-based Systems", Eindhoeven, Netherlands, 2016, https://hal.inria.fr/hal-01423494.

[25] M. BASSEUR, B. DERBEL, A. GOEFFON, A. LIEFOOGHE.*Experiments on greedy and local search heuristics for d–dimensional hypervolume subset selection*, in "Genetic and Evolutionary Computation Conference (GECCO 2016)", Denver, United States, 2016, https://hal.archives-ouvertes.fr/hal-01302283.

[26] A. BLOT, H. HOOS, L. JOURDAN, M.-É. MARMION, H. TRAUTMANN.*MO-ParamILS: A Multi-objective Automatic Algorithm Configuration Framework*, in "Learning and Intelligent Optimization", Ischia, Italy, Lecture Notes in Computer Science, May 2016, vol. 10079, p. 32-47 [*DOI :* 10.1007/978-3-319-50349-3_3], https://hal.archives-ouvertes.fr/hal-01370392.

[27] I. CHAKROUN, N. MELAB.*HB&B@GRID: An heterogeneous grid-enabled Branch and Bound algorithm*, in "2016 International Conference on High Performance Computing & Simulation (HPCS)", Innsbruck, Austria, July 2016 [*DOI :* 10.1109/HPCSIM.2016.7568403], https://hal.inria.fr/hal-01419078.

[28] B. DERBEL, A. LIEFOOGHE, Q. ZHANG, H. AGUIRRE, K. TANAKA.*Multi-objective Local Search Based on Decomposition*, in "International Conference on Parallel Problem Solving from Nature (PPSN 2016)", Edinburgh, United Kingdom, Lecture Notes in Computer Science, 2016, vol. 9921, p. 431 - 441 [*DOI :* 10.1007/978-3-319-45823-6_40], https://hal.archives-ouvertes.fr/hal-01380632.

[29] N. DUPIN, E.-G. TALBI.*Dual heuristics and new lower bounds for the challenge EURO/ROADEF*, in "Matheuristics‘2016 6th Int. Workshop on Model-based Metaheuristics", Brussels, Belgium, 2016, https://hal.inria.fr/hal-01423503.

[30] N. DUPIN, E.-G. TALBI.*Matheuristics for the discrete unit committment problem with min-stop ramping constraints*, in "Matheuristics‘2016 6th Int. Workshop on Model-based Metaheuristics", Brussels, Belgium, 2016, https://hal.inria.fr/hal-01423502.

[31] N. DUPIN, E.-G. TALBI.*Robust scheduling of nuclear power plant's maintenances*, in "MIM'2016 8th IFAC Conference on Manufacturing Modelling Management and Control", Troyes, France, 2016, https://hal.inria.fr/hal-01423499.

[32] Z. GARROUSSI, R. ELLAIA, E.-G. TALBI.*Appliance scheduling in a smart home using a multiobjective evolutionary algorithm*, in "4rd IEEE International Renewable and Sustainable Energy Conference (IRSEC)", Marrakech, Morocco, 2016, https://hal.inria.fr/hal-01423495.

[33] V. GAUTIER, M. MEZMAZ, D. TUYTTENS, N. MELAB.*Vectorization of local search for solving flow-shop scheduling problem on Xeon Phi™ MIC co-processors*, in "2016 International Conference on High Performance Computing & Simulation (HPCS)", Innsbruck, Austria, July 2016 [*DOI :* 10.1109/HPCSIM.2016.7568407], https://hal.inria.fr/hal-01419077.

[34] S. JACQUIN, E. ALLART, F. DUFOSSÉ, L. JOURDAN.*Evolutionary Algorithm for bi-objective Just-in-Time Job-shop*, in "2016 IEEE Symposium Series on Computational Intelligence (SSCI 2016)", Athens, Greece, December 2016, https://hal.archives-ouvertes.fr/hal-01420928.

[35] C. JANKEE, S. VEREL, B. DERBEL, C. FONLUPT.*A Fitness Cloud Model for Adaptive Metaheuristic Selection Methods*, in "14th International Conference on Parallel Problem Solving from Nature (PPSN2016)", Edinburgh, United Kingdom, J. HANDL, E. HART, P. R. LEWIS, M. LÓPEZ-IBÁÑEZ, G. OCHOA, B. PAECHTER (editors), Lecture Notes in Computer Science (LNCS), Springer, September 2016, vol. 9921, p. 80-90 [*DOI :* 10.1007/978-3-319-45823-6_8], https://hal.archives-ouvertes.fr/hal-01355249.

[36] M.-É. KESSACI-MARMION, H. AGUIRRE, C. DHAENENS, L. JOURDAN, K. TANAKA.*Multi-objective Neutral Neighbors? What could be the definition(s)?*, in "Genetic and Evolutionary Computation Conference", Denver, CO, United States, T. FRIEDRICH, F. NEUMANN, A. M. SUTTON (editors), Proceedings of the 2016 on Genetic and Evolutionary Computation Conference, ACM, 2016, p. 349–356 [*DOI :* 10.1145/2908812.2908902], https://hal.archives-ouvertes.fr/hal-01420900.

[37] A. LIEFOOGHE, B. DERBEL.*A Correlation Analysis of Set Quality Indicator Values in Multiobjective Optimization*, in "Genetic and Evolutionary Computation Conference (GECCO 2016)", Denver, United States, July 2016, https://hal.archives-ouvertes.fr/hal-01159961.

[38] J. PUENTE, C. R. VELA, I. GONZALEZ-RODRIGUEZ, E.-G. TALBI.*SimGA : A simulating approach to fuzzy job shop problems* , in "META'2016 6th International Conference on Metaheuristics and Nature Inspired Computing", Marrakech, Morocco, 2016, https://hal.archives-ouvertes.fr/hal-01424935.

[39] M. SAGAWA, H. AGUIRRE, F. DAOLIO, A. LIEFOOGHE, B. DERBEL, S. VEREL, K. TANAKA.*Learning variable importance to guide recombination*, in "IEEE Symposium on Computational Intelligence in Multi-criteria Decision-Making (IEEE MCDM 2016)", Athens, Greece, 2016, https://hal.archives-ouvertes.fr/hal-01380646.

[40] C. TIAGO, J. GMYS, N. MELAB, D. C. J. FRANCISCO HERON, D. TUYTTENS.*A GPU-Based Backtracking Algorithm for Permutation Combinatorial Problems*, in "16th International Conference, ICA3PP 2016", Granada, Spain, Lecture Notes in Computer Science, Springer International Publishing, December 2016, vol. 10048, 15 [*DOI :* 10.1007/978-3-319-49583-5_24], https://hal.inria.fr/hal-01419073.

[41] M. VANDROMME, J. JACQUES, J. TAILLARD, L. JOURDAN, C. DHAENENS.*A scalable biclustering method for heterogeneous medical data*, in "International Workshop on Machine Learning, Optimization and Big

Data", Volterra, Italy, Lecture Notes in Computer Science, August 2016, vol. 10122, 12, https://hal.inria.fr/hal-01420947.

Conferences without Proceedings

[42] A.-L. BEDENEL, C. BIERNACKI, L. JOURDAN.*Matching of descriptors evolving over time : Application to online insurance comparison* , in "48èmes Journées des Statistiques Française", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01381766.

[43] S. DUFOURNY, C. DHAENENS.*Optimization of business game decisions*, in "IFAC / IEEE / IFORS / IFIP / INFORMS MIM 2016 Conference on Manufacturing Modelling, Management, and Control", Troyes, France, June 2016, https://hal.archives-ouvertes.fr/hal-01347368.

[44] L. MOUSIN, L. JOURDAN, M.-É. MARMION, C. DHAENENS.*Feature Selection using Tabu Search with Learning Memory: Learning Tabu Search*, in "Learning and Intelligent OptimizatioN Conference LION 10", Ischia Island (Napoli), Italy, May 2016, https://hal.archives-ouvertes.fr/hal-01370396.

Scientific Books (or Scientific Book chapters)

[45] C. DHAENENS, L. JOURDAN.*Metaheuristics for Big Data*, Wiley-ISTE, August 2016, 212, https://hal.inria.fr/hal-01418464.

Books or Proceedings Editing

[46] M.-J. BLESA, C. BLUM, A. CANGELOSI, V. CUTELLO, A. DI NUOVO, M. PAVONE, E.-G. TALBI (editors). *Hybrid metaheuristics*, Lecture Notes in Computer Science LNCS No.9668, Springer, 2016, https://hal.inria.fr/hal-01423491.

Other Publications

[47] D. BROCKHOFF, T. TUSAR, D. TUSAR, T. WAGNER, N. HANSEN, A. AUGER.*Biobjective Performance Assessment with the COCO Platform*, May 2016, ArXiv e-prints, arXiv:1605.01746, https://hal.inria.fr/hal-01315317.

[48] N. HANSEN, A. AUGER, D. BROCKHOFF, D. TUSAR, T. TUSAR.*COCO: Performance Assessment*, May 2016, ArXiv e-prints, arXiv:1605.03560, https://hal.inria.fr/hal-01315318.

[49] N. HANSEN, A. AUGER, O. MERSMANN, T. TUSAR, D. BROCKHOFF.*COCO: A Platform for Comparing Continuous Optimizers in a Black-Box Setting*, July 2016, ArXiv e-prints, arXiv:1603.08785, https://hal.inria.fr/hal-01294124.

References in notes

[50] C. A. COELLO COELLO, D. A. VAN VELDHUIZEN, G. B. LAMONT (editors). *Evolutionary algorithms for solving multi-objective problems*, Kluwer Academic Press, 2002.

[51] M. BASSEUR.*Design of cooperative algorithms for multi-objective optimization: Application to the Flow-shop scheduling problem*, University of Sciences and Technology of Lille, France, June 2005.

[52] C. COTTA, E.-G. TALBI, E. ALBA.*Parallel hybrid approaches*, in "Parallel Metaheuristics", USA, J. Wiley and Sons, 2005, p. 347–370.

[53] K. DEB.*Multi-objective optimization using evolutionary algorithms*, John Wiley and sons, 2001.

[54] D. E. GOLDBERG.*Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Massachusett, USA, 1989.

[55] A. J. NEBRO, F. LUNA, E.-G. TALBI, E. ALBA.*Parallel multi-objective optimization*, in "Parallel Metaheuristics", USA, J. Wiley and Sons, 2005, p. 371–394.

[56] E.-G. TALBI.*A Taxonomy of Hybrid Metaheuristics*, in "Journal of Heuristics", 2002, vol. 8, n^{o} 5, p. 541–564.

Project-Team DREAMPAL

Dynamic Reconfigurable Massively Parallel Architectures and Languages

IN PARTNERSHIP WITH:

CNRS

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER

Lille - Nord Europe

THEME

Architecture, Languages and Compilation

Table of contents

Project-Team DREAMPAL

Creation of the Team: 2013 January 01, updated into Project-Team: 2015 January 01, end of the Project-Team: 2016 December 31

Keywords:

Computer Science and Digital Science:

1.1.2. - Hardware accelerators (GPGPU, FPGA, etc.)
1.1.12. - Non-conventional architectures
2.1.1. - Semantics of programming languages
2.4.2. - Model-checking

Other Research Topics and Application Domains:

6.6. - Embedded systems
7.2.1. - Smart vehicles

1. Members

Research Scientist
Vlad Rusu [Team leader, Inria, Researcher, HDR]

Faculty Members
Ahmad Shadi Aljendi [Univ. Lille I, Teaching Assistant, until Sep 2016]
Rabie Ben Atitallah [Univ. Valenciennes, Associate Professor, HDR]
Jean-Luc Dekeyser [Univ. Lille I, Professor, HDR]
Frederic Guyomarch [Univ. Lille I, Associate Professor]
Philippe Marquet [Univ. Lille I, Associate Professor]
Samy Meftali [Univ. Lille I, Associate Professor]

PhD Students
Karim Ali [Univ. Valenciennes, until Sep 2016]
Wissem Chouchene [Univ. Lille I]

Post-Doctoral Fellow
Andrei Arusoaie [Inria, until Feb 2016]

Administrative Assistant
Corinne Jamroz [Inria]

2. Overall Objectives

2.1. Executive Summary

Standard Integrated Circuits are reaching their limits and need to evolve in order to meet the requirements of next-generation computing. We anticipate that FPGAs (Field Programmable Gate Arrays) will play a major role in this evolution: FPGAs are currently only one or two generations behind the most advanced technologies for standard processors, and their application-specific hardware is an order of magnitude faster than software solutions on standard processors. One of the most promising evolutions are next-generation 3D-FPGAs, which, thanks to their fast reconfiguration and inherent parallelism, will enable users to build dynamically reconfigurable, massively parallel hardware architectures around them. This new paradigm opens many opportunities for research, since, to our best knowledge, there are no methodologies for building such architectures, and there are no dedicated languages for programming on them.

We shall thus address the following topics: proposing an execution model and a design environment, in which users can build customized massively parallel dynamically reconfigurable hardware architectures, benefiting from the reconfiguration speed and parallelism of 3D-FPGAs; proposing dedicated languages for programming applications on such architectures; and designing software engineering tools for those languages: compilers, simulators, and formal verifiers. The overall objective is to enable an efficient and safe programming on the customized architectures. Our target application domain are embedded systems performing intensive signal/image processing (e.g., smart cameras, radars, and set-top boxes)

3. Research Program

3.1. New Models for New Technologies

Over the past 25 years there have been several hardware-architecture generations dedicated to massively parallel computing. We have contributed to them in the past, and shall continue doing so in the Dreampal project. The three generations, chronologically ordered, are:

- Supercomputers from the 80s and 90s, based on massively parallel architectures that are more or less distributed (from the Cray T3D or Connection Machine CM2 to GRID 5000). Computer scientists have proposed methods and tools for mapping sequential algorithms to those parallel architectures in order to extract maximum power from them. We have contributed in this area in the past.
- Parallelism pervades the chips! A new challenge appears: hardware/software co-design, in order to obtain performance gains by designing algorithms together with the parallel architectures of chips adapted to the algorithms. During the previous decade many studies, including ours in the Inria DaRT team, were dedicated to this type of co-design. DaRT has contributed to the development of the OMG MARTE standard (http://www.omgmarte.org) and to its implementation on several parallel platforms. Gaspard2, our implementation of this concept, was identified as one of the key software tools developed at Inria: http://www.inria.fr/en/centre/lille/research/platforms-and-flagship-software/flagship-software.
- The new challenge of the 2010s is, in our opinion, the integration of dynamic reconfiguration and massive parallelism. New circuits with high-density integration and supporting dynamic hardware reconfiguration have been proposed. In such architectures one can dynamically change the architecture while an algorithm is running on it. The Dynamic Partial Reconfiguration (DPR) feature offered by recent FPGA boards even allows, in theory, to generate optimized hardware at runtime, by adding, removing, and replacing components on a by-need basis. This integration of dynamic reconfiguration and massive parallelism induces a new degree of complexity, which we, as computer scientists, need to understand and deal with order to make possible the design of applications running on such architectures. This is the main challenge that we address in the Dreampal project. We note that we adress these problems as computer scientists; we do, however, collaborate with electronics specialists in order to benefit from their expertise in 3-D FPGAs.

Excerpt from the HiPEAC vision 2011/12

> *"The advent of 3D stacking enables higher levels of integration and reduced costs for off-chip communications. The overall complexity is managed due to the separation in different dies, independently designed."*

FPGAs (Field Programmable Gate Arrays) are configurable circuits that have emerged as a privileged target platform for intensive signal processing applications. FPGAs take advantage of the latest technological developments in circuits. For example, the Virtex7 from Xilinx offers a 28-nanometer integration, which is only one or two generations behind the latest general-purpose processors. 3D-Stacked Integrated Circuits (3D SICs) consist of two or more conventional 2D circuits stacked on the top of each other and built into the same IC. Recently, 3D SICs have been released by Xilinx for the Virtex 7 FPGA family. 3D integration will vastly increase the integration capabilities of FPGA circuits. The convergence of massive parallelism and dynamic reconfiguration in inevitable: we believe it is one of the main challenges in computing for the current decade.

By incorporating the configuration and/or data/program memory on the top of the FPGA fabric, with fast and numerous connections between memory and elementary logic blocks ($\sim$10000 connections between dies), it will be possible to obtain dynamically reconfigurable computing platforms with a very high reconfiguration rate. Such a rate was not possible before, due to the serial nature of the interface between the configuration memory and the FPGA fabric itself. The FPGA technology also enables massively parallel architectures due to the large number of programmable logic fabrics available on the chip. For instance, Xilinx demonstrated 3600 8-bit picoBlaze softcore processors running simultaneously on the Virtex-7 2000T FPGA. For specific applications, picoBlaze can be replaced by specialized hardware accelerators or other IPs (Intellectual Property) components. This opens the possibility of creating massively parallel IP-based machines.

3.2. Multi-softcore on 3D FPGA

From the 2010 Xilinx white paper on FPGAs:

> *“Unlike a processor, in which architecture of the ALU is fixed and designed in a general-purpose manner to execute various operations, the CLBs (configurable logic blocks) can be programmed with just the operations needed by the application... The FPGA architecture provides the flexibility to create a massive array of application-specific ALUs..The new solution enables high-bandwidth connectivity between multiple die by providing a much greater number of connections... enabling the integration of massive quantities of interconnect logic resources within a single package”*

Softcore processors are processors implemented using hardware synthesis. Proprietary solutions include PicoBlaze, MicroBlaze, Nios, and Nios II; open-source solutions include Leon, OpenRisk, and FC16. The choice is wide and many new solutions emerge, including multi-softcore implementations on FPGAs. An alternative to softcores are hardware accelerators on FPGAs, which are dedicated circuits that are an order of magnitude faster than softcores. Between these two approaches, there are other various approaches that connect IPs to softcores, in which, the processor’s machine-code language is extended, and IP invocations become new instructions. We envisage a new class of softcores (we call them reflective softcores [0]), where almost everything is implemented in IPs; only the control flow is assigned to the softcore itself. The partial dynamic reconfiguration of next-generation FPGAs makes such dynamic IP management possible in practice. We believe that efficient reflective softcores on the new 3D-FPGAs should be as small as possible: low-performance generic hardware components (ALU, registers, memory, I/O...) should be replaced by dedicated high-performance IPs.

We are developing a sofcore processor called HoMade (http://www.lifl.fr/~dekeyser/Homade) following these ideas.

In the multi-reflective softcores that we develop, some softcores will be slaves and others will be masters. Massively parallel dynamically reconfigurable architectures of softcores can thus be envisaged. This requires, additionally, a parallel management of the partial dynamic reconfiguration system. This can be done, for example, on a given subset of softcores: a massively parallel reconfiguration will replace the current replication of a given IP with the replication of a new IP. Thanks to the new 3D-FPGAs this task can be performed efficiently and in parallel using the large number of 3D communication links (Through-Silicon-Vias). Our roadmap for HoMade is to evolve towards this multi-reflective softcore model.

3.3. When Hardware Meets Software

HIPEAC vision 2011/12: *"The number of cores and instruction set extensions increases with every new generation, requiring changes in the software to effectively exploit the new features."*

[0] Hereafter, by reflective system, we mean a system that is able to modify its own structure and behaviour while it is running. A reflective softcore thus dynamically adds, removes, and replaces IPs in the application running on it, and is able to dynamically modify its own program memory, thereby dynamically altering the program it is executing.

When the new massively parallel dynamically reconfigurable architectures become reality users will need languages for programming software applications on them. The languages will be themselves dynamic and parallel, in order to reflect and to fully exploit the dynamicity and parallelism of the architectures. Thus, developers will be able to invoke reconfiguration and call parallel instructions in their programs. This expressiveness comes with a cost, however, because new classes of bugs can be induced by the interaction between dynamic reconfiguration and parallelism; for example, deadlocks due to waiting for output from an IP that does not exist any more due to a reconfiguration. The detection and elimination of such bugs before deployment is paramount for cost-effectiveness and safety reasons.

Thus, we shall build an environment for developing software on parallel, dynamically reconfigurable architectures that will include languages and adequate formal analyses and verification tools for them, in addition to more traditional tools (emulators, compilers, etc). To this end we shall be using formal-semantics frameworks associated with easy-to-use formal verification tools in order to formally define our languages of interest and allow users to formally verify their programs. The K semantic framework (http://k-framework.org), developed jointly by Univs. Urbana Champaign, USA, and Iasi, Romania) is one such framework, which is mature enough (it has allowed defining a formal semantics of the largest subset of the C language to date, as well as many other languages from essentially all programming paradigms) and is familiar to us from previous work. In K, one can rapidly prototype a language definition and try several versions of the syntax and semantics of instructions. This is important in our project, where the proposed programming languages (in particular, the HoMade assembly language) will go through several versions before being stabilized. Moreover, once a language is defined in K one gets an interpreter of the language and one gains access to formal verification tools for free. We are also developing new analysis verification tools for K (in collaboration with the K team), which will be adapted and used in the Dreampal project.

4. Highlights of the Year

4.1. Highlights of the Year

2016 is the last year of Dreampal's existence as an Inria project-team. Due to different scientific objectives, three of the members (S. Meftali, J.L. Dekeyser, P. Marquet) will create a group within the Cristal laboratory, while the team leader V. Rusu will collaborate with the 2xs team within Cristal. Frédéric Guyomarch joined the L2EP laboratory, and external collaborator Rabie Ben Atitallah continues his activity in the LAMIH laboratory in Valenciennes.

This activity report has been written by the team leader, based on the information available to him at the time of its writing. Any activity, e.g., by other team members, not reflected in the report, is only missing because of lack of input from the people concerned.

5. New Software and Platforms

5.1. HoMade

KEYWORDS: SoC - Multicore - Softcore
FUNCTIONAL DESCRIPTION

HoMade is a softcore processor. The current version is reflective (i.e., the program it executes is self-modifiable), and statically configurable, dynamically reconfigurable multi-processors are the next steps. Users have to add to it the functionality they need in their applications via IPs. We have also being developing a library of IPs for the most common processor functions (ALU, registers, ...). All the design is in VHDL except for some schematic specifications.

- Participant: Jean Luc Dekeyser
- Partner: LIFL
- Contact: Jean Luc Dekeyser
- URL: https://sites.google.com/site/homadeguideen/home

5.2. JHomade

FUNCTIONAL DESCRIPTION

JHomade is a software suite written in JAVA, including compilers and tools for the HoMade processor. It allows us to compile HiHope programs to Homade machine code and load the resulting binaries on FPGA boards. It was first released in 2013. The second version in 2014 includes several new features, like a C-frontend, a binary decoder and a code-generator for VHDL simulation. New features of the HiHope language are described in more detail in Section.

- Contact: Vlad Rusu
- URL: https://gforge.inria.fr/frs/?group_id=3646

6. New Results

6.1. A Language-Independent Proof System for Full Program Equivalence

Two programs are mutually equivalent if, for the same input, either they both diverge or they both terminate with the same result. Mutual equivalence is an adequate notion of equivalence for programs written in deterministic languages. It is useful in many contexts, such as capturing the correctness of program transformations within the same language, or capturing the correctness of compilers between two different languages. In [11] we introduce a language-independent proof system for mutual equivalence, which is para-metric in the operational semantics of two languages and in a state-similarity relation. The proof system is sound: if it terminates then it establishes the mutual equivalence of the programs given to it as input. We illustrate it on two programs in two different languages (an imperative one and a functional one), that both compute the Collatz sequence. The Collatz sequence is an interesting case study since it is not known wether the sequence terminates or not; nevertheless, our proof system shows that the two programs are mutually equivalent (even if we cannot establish termination or divergence of either one).

6.2. A Generic Framework for Symbolic Execution: a Coinductive Approach

In [12] we propose a language-independent symbolic execution framework. The approach is parameterised by a language definition, which consists of a signature for the lan-guage's syntax and execution infrastructure, a model interpreting the signature, and rewrite rules for the language's operational semantics. Then, symbolic execution amounts to computing symbolic paths using a derivative operation. We prove that the symbolic execution thus defined has the properties naturally expected from it, meaning that the feasible symbolic executions of a program and the concrete executions of the same program mutually simulate each other. We also show how a coinduction-based extension of symbolic execution can be used for the deductive verification of programs. We show how the proposed symbolic-execution approach, and the coinductive verification technique based on it, can be seamlessly implemented in language definition frameworks based on rewriting such as the K framework. A prototype implementation of our approach has been developed in K. We illustrate it on the symbolic analysis and deductive verification of nontrivial programs.

6.3. Circuit Merging versus Dynamic Partial Reconfiguration -The HoMade Implementation

One goal of reconfiguration is to save power and occupied resources. In [13] we compare two different kinds of reconfiguration available on field-programmable gate arrays (FPGA) and we discuss their pros and cons. The first method that we study is circuit merging. This type of reconfiguration methods consists in sharing common resources between different circuits. The second method that we explore is dynamic partial reconfiguration (DPR). It is specific to some FPGA, allowing well defined reconfigurable parts to be modified during run-time. We show that DPR, when available, has good and more predictable result in terms of occupied area. There is still a huge overhead in term of time and power consumption during the reconfiguration phase.

Therefore we show that circuit merging remains an interesting solution on FPGA because it is not vendor specific and the reconfiguration time is around a clock cycle. Besides, good merging algorithms exist eventhough FPGA physical synthesis flow makes it hard to predict the real performance of the merged circuit during the optimization. We establish our comparison in the context of the HoMade processo

6.4. Language Definitions as Rewrite Theories

K is a formal framework for defining operational semantics of programming languages. The K-Maude compiler translates K language definitions to Maude rewrite theories. The compiler enables program execution by using the Maude rewrite engine with the compiled definitions, and program analysis by using various Maude analysis tools. K supports symbolic execution in Maude by means of an automatic transformation of language definitions. The transformed definition is called the symbolic extension of the original definnition. In [14] we investigate the theoretical relationship between K language definitions and their Maude translations, between symbolic extensions of K definitions and their Maude translations, and how the relationship between K definitions and their symbolic extensions is reflected on their respective representations in Maude. In particular, the results show how analysis performed with Maude tools can be formally lifted up to the original language definitions.

6.5. SCAC-Net: Reconfigurable Interconnection Network in SCAC Massively parallel SoC

Parallel communication plays a critical role in massively parallel systems, especially in distributed memory systems executing parallel programs on shared data. Therefore, integrating an interconnection network in these systems becomes essential to ensure data inter-nodes exchange. Choosing the most effective communication structure must meet certain criteria: speed, size and power consumption. Indeed, the communication phase should be as fast as possible to avoid compromising parallel computing, using small and low power consumption modules to facilitate the interconnection network extensibility in a scalable system. To meet these criteria and based on a module reuse methodology, we chose to integrate a reconfigurable SCAC-Net interconnection network to communicate data in SCAC Massively parallel SoC. In [15] we present the detailed hardware implementation and discusses the performance evaluation of the proposed reconfigurable SCAC-Net network.

6.6. Proving Reachability-Logic Formulas Incrementally

Reachability Logic (RL) is a formalism for defining the operational semantics of programming languages and for specifying program properties. As a program logic it can be seen as a language-independent alternative to Hoare Logics. Several verification techniques have been proposed for RL, all of which have a circular nature: the RL formula under proof can circularly be used as a hypothesis in the proof of another RL formula, or even in its own proof. This feature is essential for dealing with possibly unbounded repetitive behaviour (e.g., program loops). The downside of such approaches is that the verification of a set of RL formulas is monolithic, i.e., either all formulas in the set are proved valid, or nothing can be inferred about any of the formula's validity or invalidity. In [16] we propose a new, incremental method for proving a large class of RL formulas. The proposed method takes as input a given RL formula under proof (corresponding to a given program fragment), together with a (possibly empty) set of other valid RL formulas (e.g., already proved using our method), which specify sub-programs of the program fragment under verification. It then checks certain conditions are shown to be equivalent to the validity of the RL formula under proof. A newly proved formula can then be incrementally used in the proof of other RL formulas, corresponding to larger program fragments. The process is repeated until the whole program is proved. We illustrate our approach by verifying the nontrivial Knuth-Morris-Pratt string-matching program.

7. Partnerships and Cooperations

7.1. International Initiatives

7.1.1. Inria International Partners

7.1.1.1. Informal International Partners

In 2016 we have continued our strong and long-term collaboration with Prof. Dorel Lucanu's group Univ. Iasi as witnessed by the co-authored publications (3 journals and 1 conference). Vlad Rusu serves as "external advisor for PhD students" in Prof. Lucanu's group. In 2016 we have also had notable interactions with Prof. José Meseguer (Univ. Illinois at Urbana Champaign, USA), which consisted in sharing ideas and mutual reading and commenting advanced drafts prior to submission in journals/conferences.

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Selection

8.1.1.1. Member of the Conference Program Committees

V.Rusu was a member in the PC of the 2016 edition of the Int. Workshop on Rewriting Logic and Applications, and will be the organizer of the next edition of the event. He was was also PC member of Approches Formelles pour la Validation Logicielle (AFADL'2016).

8.1.2. Research Administration

Vlad Rusu is elected member at Inria's Evaluation Committee. As such he has been involved in several activities regarding promotions of researchers, team creations, and team evaluations.

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Licence : V.Rusu, Logic, 30hrs, L3, Univ. Lille, France.

Doctorat : V. Rusu, External adviser for PhD students, Univ. Iasi, Romania.

9. Bibliography

Major publications by the team in recent years

[1] A. AIT EL CADI, O. SOUISSI, R. BEN ATITALLAH, N. BELANGER, A. ARTIBA.*Mathematical programming models for scheduling in a CPU/FPGA architecture with heterogeneous communication delays*, in "Journal of Intelligent Manufacturing", April 2015, p. 1-12 [*DOI :* 10.1007/S10845-015-1075-Z], https://hal.inria.fr/hal-01247399.

[2] K. M. A. ALI, R. B. ATITALLAH, N. FAKHFAKH, J.-L. DEKEYSER.*Using hardware parallelism for reducing power consumption in video streaming applications*, in "10th International Symposium on Reconfigurable Communication-centric Systems-on-Chip (ReCoSoC), 2015", Bremen, Germany, June 2015 [*DOI :* 10.1109/RECOSOC.2015.7238104], https://hal.inria.fr/hal-01247130.

[3] M. BAKLOUTI, P. MARQUET, J.-L. DEKEYSER, M. ABID.*FPGA-based many-core System-on-Chip design*, in "Microprocessors and Microsystems", 2015, 38 [*DOI :* 10.1016/J.MICPRO.2015.03.007], https://hal.inria.fr/hal-01144977.

[4] S. CIOBACA, D. LUCANU, V. RUSU, G. ROSU.*A Language-Independent Proof System for Full Program Equivalence*, in "Formal Aspects of Computing", 2016, vol. 28, n^{o} 3, p. 469–497 [*DOI :* 10.1007/S00165-016-0361-7], https://hal.inria.fr/hal-01245528.

[5] D. LUCANU, V. RUSU, A. ARUSOAIE.*A Generic Framework for Symbolic Execution: a Coinductive Approach*, in "Journal of Symbolic Computation", 2016, Accepted for publication [*DOI :* 10.1016/J.JSC.2016.07.012], https://hal.inria.fr/hal-01238696.

[6] D. LUCANU, V. RUSU, A. ARUSOAIE, D. NOWAK.*Verifying Reachability-Logic Properties on Rewriting-Logic Specifications*, in "Logic, Rewriting, and Concurrency - Festschrift Symposium in Honor of José Meseguer", Urbana Champaign, United States, Logic, Rewriting, and Concurrency Essays Dedicated to José Meseguer on the Occasion of His 65th Birthday, Springer Verlag, September 2015, vol. 9200 [*DOI :* 10.1007/978-3-319-23165-5_21], https://hal.inria.fr/hal-01158941.

[7] D. LUCANU, V. RUSU.*Program Equivalence by Circular Reasoning*, in "Formal Aspects of Computing", 2015, vol. 27, n^{o} 4, p. 701-726 [*DOI :* 10.1007/S00165-014-0319-6], https://hal.inria.fr/hal-01065830.

[8] V. RUSU, D. LUCANU, T.-F. ŞERBĂNUŢĂ, A. ARUSOAIE, A. ŞTEFĂNESCU, G. ROŞU.*Language Definitions as Rewrite Theories*, in "Journal of Logic and Algebraic Methods in Programming", 2016, vol. 85, n^{o} 1, p. 98–120 [*DOI :* 10.1016/J.JLAMP.2015.09.001], https://hal.inria.fr/hal-01186005.

[9] V. VISWANATHAN, R. B. ATITALLAH, J.-L. DEKEYSER.*Massively Parallel Dynamically Reconfigurable Multi-FPGA Computing System* , in "IEEE 23rd Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM), 2015", Vancouver, BC , Canada, May 2015 [*DOI :* 10.1109/FCCM.2015.13], https://hal.inria.fr/hal-01247116.

[10] V. VISWANATHAN, R. BEN ATITALLAH, J.-L. DEKEYSER, B. NAKACHE, M. NAKACHE.*A Parallel And Scalable Multi-FPGA based Architecture for High Performance Applications*, in "The 2015 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, FPGA '15", Monterey, California, United States, February 2015 [*DOI :* 10.1145/2684746.2689115], https://hal.inria.fr/hal-01247133.

Publications of the year

Articles in International Peer-Reviewed Journal

[11] S. CIOBACA, D. LUCANU, V. RUSU, G. ROSU.*A Language-Independent Proof System for Full Program Equivalence*, in "Formal Aspects of Computing", 2016, vol. 28, n^{o} 3, p. 469–497 [*DOI :* 10.1007/S00165-016-0361-7], https://hal.inria.fr/hal-01245528.

[12] D. LUCANU, V. RUSU, A. ARUSOAIE.*A Generic Framework for Symbolic Execution: a Coinductive Approach*, in "Journal of Symbolic Computation", 2016, Accepted for publication [*DOI :* 10.1016/J.JSC.2016.07.012], https://hal.inria.fr/hal-01238696.

[13] J. PERIER, W. CHOUCHENE, J.-L. DEKEYSER.*Circuit Merging versus Dynamic Partial Reconfiguration -The HoMade Implementation*, in "i-manager's Journal on Embedded Systems(JES)", 2016, https://hal.inria.fr/hal-01245800.

[14] V. RUSU, D. LUCANU, T.-F. ŞERBĂNUŢĂ, A. ARUSOAIE, A. ŞTEFĂNESCU, G. ROŞU.*Language Definitions as Rewrite Theories*, in "Journal of Logic and Algebraic Methods in Programming", 2016, vol. 85, n° 1, p. 98–120 [*DOI :* 10.1016/J.JLAMP.2015.09.001], https://hal.inria.fr/hal-01186005.

International Conferences with Proceedings

[15] H. KRICHENE, M. BAKLOUTI, M. ABID, P. MARQUET, J.-L. DEKEYSER, S. MEFTALI.*SCAC-Net: Reconfigurable Interconnection Network in SCAC Massively parallel SoC*, in "The 24th Euromicro International Conference on Parallel, Distributed and Network-Based Processing", Heraklion, Greece, February 2016, https://hal.inria.fr/hal-01247298.

[16] V. RUSU, A. ARUSOAIE.*Proving Reachability-Logic Formulas Incrementally*, in "11th International Workshop on Rewriting Logic and its Applications", Eindhoven, Netherlands, April 2016, https://hal.inria.fr/hal-01282379.

Project-Team FUN

self-organizing Future Ubiquitous Network

RESEARCH CENTER
Lille - Nord Europe

THEME
Networks and Telecommunications

Table of contents

Project-Team FUN

Creation of the Team: 2012 January 01, updated into Project-Team: 2013 July 01

Keywords:

Computer Science and Digital Science:

1.2.1. - Dynamic reconfiguration
1.2.3. - Routing
1.2.4. - QoS, performance evaluation
1.2.5. - Internet of things
1.2.6. - Sensor networks
1.2.7. - Cyber-physical systems
1.4. - Ubiquitous Systems
5.10.6. - Swarm robotics

Other Research Topics and Application Domains:

5.1. - Factory of the future
5.6. - Robotic systems
5.9. - Industrial maintenance
6.4. - Internet of things
7. - Transport and logistics
8. - Smart Cities and Territories

1. Members

Research Scientists

Nathalie Mitton [Team leader, Inria, Senior Researcher, HDR]
Cristina Cano Bastidas [Inria, Researcher, until Jul 2016]
Simon Duquennoy [Inria, Starting Research position]
Valeria Loscri [Inria, Senior Researcher]
Tahiry Razafindralambo [Inria, Senior Researcher, until Jul 2016, HDR]

Technical Staff

Emilio Compagnone [Inria, until Aug 2016]
Clement Fumey [Inria, until Aug 2016]
Salvatore Guzzo Bonifacio [Inria, until Feb 2016]
Julien Vandaele [Inria]
Gabriele Sabatino [Inria]

PhD Students

Mouna Masmoudi [Inria, until Jul 2016]
Abdoul Aziz Mbacke [Inria]
Jad Nassar [HEI]
Riccardo Petrolo [Inria, until Oct 2016]
Jean Cristanel Razafimandimby Anjalalaina [Inria]
Viktor Toldov [Univ. Lille I, until Dec 2016]

Post-Doctoral Fellow

Farouk Mezghani [Inria, from Sep 2016]

Visiting Scientists

Roberto Morabito [Ericsson, from Sep 2016]
William Pretorius [Stellenbosch Univ, from Sep 2016 until Nov 2016]

Administrative Assistant

Anne Rejl [Inria]

Others

Alexandre Peyrard [Inria, Internship, from Mar 2016 until Aug 2016]
Jonathan Prin [Inria, Internship, from Jun 2016 until Jul 2016]
Rahul Vyas [Inria, Internship, from Apr 2016 until Jun 2016]

2. Overall Objectives

2.1. Overall Objectives

Context.

The Internet of Things [62] is a large concept with multiple definitions. However, the main concepts are the same in every vision and could be summed up as follows: *Imagine a world where every object has the capacity to communicate with its environment. Everything can be both analogue and digitally approached - reformulates our relationship with objects - things - as well as the objects themselves. Any object relates not only to you, but also to other objects, relations or values in a database. In this world, you are no longer alone, anywhere.* (Internet of Things council).

Future Ubiquitous Networks (FUN) are part of the Internet of Things. They are composed of tens to thousands heterogeneous hardware-constrained devices that interact with our environment and the physical world. These devices have limited resources in terms of storage and computing capacities and energy. They communicate through unreliable and unpredictable short-range wireless links and run on batteries that are not envisaged to be changed in current systems since generally deployed in hostile environments. Providing FUNs with energy saving protocols is thus a key issue. Due to these specific features, any centralized control is not conceivable, the new generation of FUNs must be autonomous, be self-organized and dynamically adapt to their environment. The devices that compose CPNs can be sensors, small robots, RFID readers or tags.

Objects or things can now communicate with their environment through the use for instance of an RFID (Radio Frequency IDentification) tag that provides them a unique identifier (ID) and a way to communicate through radio waves.

In the case of a simple passive **RFID tag**, the thing only embeds a tag equipped with an antenna and some memory. To communicate, it needs to be powered by the electromagnetic field of an RFID reader. This reader may then broadcast the information read on tag over a network.

When this tag is equipped with a battery, it is now able to communicate with nearby things similar to itself that may relay its message. Tags can also be equipped with additional capacity and sensors (for light, temperature, etc...). The Internet of Things can thus now refer to a **wireless sensor** network in which each sensor sends the data it collects over its environment and then sends it to a sink, *i.e.* a special sensor node able to analyze those data. In every case, RFID tags or sensor nodes can **be moved unexpectedly** like hold by moving things or animals. We speak then about **'*undergone* mobility'**.

So far, things can thus communicate information about their environment. But when the capacity of sensors is extended even further, they can also act on their environment (for instance, the detection of an event (fire) may trigger an action like switching the light or fire hoses on). Sensor nodes become **actuators**. When this extended capacity is the faculty to move, actuators are also referred as actors or robots. In this latter case, the mobility is computed on purpose, we then speak about **'*controlled* mobility'**. Actuators are not moved but move by themselves.

The FUN research group aims to focus on self-organizing techniques for these heterogeneous Future Ubiquitous Networks (FUNs). FUNs need various self-organization techniques to work properly. Self-organization encompasses neighbor discovery (which what other devices a sensor/actuator can communicate directly?), communication, self-deployment, self-localization, activity scheduling (when to wake up, when to send data to save energy without being detrimental to the well behavior of the network, etc)...

Solutions provided by FUN should facilitate the use of FUNs and rub away heterogeneity and difficulties. These techniques should be **scalable**, **energy-aware**, **standard-compliant**, should manage undergone **mobility** and take advantage of controlled mobility when available [72].

Solutions provided by FUN will consider vagaries of the realistic wireless environment by integrating cross-layer techniques in their design.

Motivation.

To date, many self-organizing techniques for wireless sensor networks and mobile ad hoc networks arise in the literature and also from the POPS research group. Some of them are very efficient for routing [64], [61], discovering neighborhood [69], [68], scheduling activity and coverage [66], localizing [73], [60], etc. Nevertheless, to the best of our knowledge, most of them **have not been validated by experimentation**, only by simulation and thus cannot consider the real impact of the wireless links and real **node mobility** in different environments. In addition, some of them rely on assumptions that are known not to be true in realistic networks such as the fact that the transmission range of a node is a perfect disk. Other may perform well only when nodes are static. None of them considers to **take advantage of controlled mobility** to enhance performances. Similarly, many propositions arise regarding self-organization in RFID networks, mainly at the middleware level [76], [65] and at the MAC layer level [58]. Although these latter propositions are generally experimented, they are validated only in static environments with very few tags and readers. To fit realistic features, such algorithms should also be evaluated with regards to scalability and mobility.

RFID and sensor/actor technologies **have not been merged**. Though, RFID readers may now be mobile and communicate in a wireless peer-to-peer manner either with other RFID readers or wireless sensor nodes and all belong to the same network. This implies a study of the standards to allow inter-dependencies in a transparent manner. Although such works have been initiated inside EPC Global working groups, research actions remain scarce.

FUN research group aims at **filling this scientific gap** by proposing self-stabilizing solutions, considering vagaries of wireless links, node mobility and heterogeneity of nodes in compliance with current standards. Validation by experimentation is mandatory to prove the effectiveness of proposed techniques in realistic environments.

FUN will investigate new protocols and communication paradigms that allow the **transparent merging** of technologies. Objects and events might interconnect while **respecting on-going standards** and building an autonomic and smart network while being compliant with hardware resources and environment. FUN expects to rub away the difficulty of use and programmability of such networks by unifying the different technologies. In addition, FUN does not only expect to validate the proposed solutions through experimentation but also to learn from these experiments and from the observation of the impact of the wireless environment to take these features into consideration in the design of future solutions.

3. Research Program

3.1. Introduction

We will focus on wireless ubiquitous networks that rely on constrained devices, i.e. with limited resources in terms of storage and computing capacities. They can be sensors, small robots, RFID readers or tags. A wireless sensor retrieves a physical measure such as light. A wireless robot is a wireless sensor that in addition has the ability to move by itself in a controlled way. A drone is a robot with the ability to manoeuvre in 3D (in the air or in the water). RFID tags are passive items that embed a unique identifier for a place or an object allowing accurate traceability. They can communicate only in the vicinity of an RFID reader. An RFID reader can be seen as a special kind of sensor in the network which data is the one read on tags. These devices may run on batteries that are not envisaged to be changed or recharged. These networks may be composed of ten to thousands of such heterogeneous devices for which energy is a key issue.

Today, most of these networks are homogeneous, i.e. composed of only one kind of devices. They have mainly been studied in application and technology silos. Because of this, they are approaching fundamental limitations especially in terms of topology deployment, management and communications, while exploiting the complementarity of heterogeneous devices and communication technologies would enlarge their capacities and the set of applications. Finally, these networks must work efficiently even in dynamic and realistic situations, i.e. they must consider by design the different dynamic parameters and automatically self-adapt to their variations.

Our overall goal is represented by Figure 1. We will investigate wireless ubiquitous IoT services for constrained devices by smartly combining **different frequency bands** and **different medium access and routing techniques**over **heterogeneous devices** in a **distributed** and **opportunistic** fashion. Our approach will always deal with **hardware constraints** and take care of **security** and **energy** issues to provide protocols that ride on **synergy** and **self-organization** between devices.

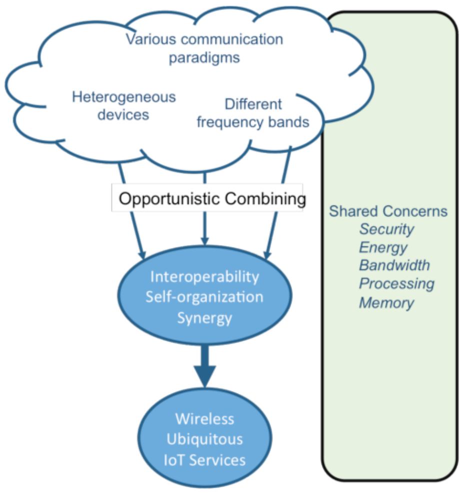


Figure 1. FUN's overal goal.

The goal of the FUN project team is to provide these next generation networks with a set of innovative and distributed self-organizing cooperative protocols to raise them to a new level of scalability, autonomy,

adaptability, manageability and performance. We aim to break these silos to exploit the full synergy between devices, making them cooperate in a single holistic network. We will consider them as networks of heterogeneous devices rather than a collection of heterogeneous networks.

To realize the full potential of these ubiquitous networks, there is a need to provide them with a set of tools that allow them to *(i)* (self-)deploy, *(ii)* self-organize, *(iii)* discover and locate each other, resources and services and *(iv)* communicate. These tools will be the basics for enabling cooperation, co-existence and witnessing a global efficient behavior. The deployment of these mechanisms is challenging since it should be achieved in spite of several limitations. The main difficulties are to provide such protocols in a **secured** and **energy-efficient** fashion in spite of :

- dynamic topology changes due to various factors such as the unreliability of the wireless medium, the wireless interferences between devices, node mobility and energy saving mechanisms;
- hardware constraints in terms of CPU and memory capacities that limit the operations and data each node can perform/collect;
- lacks of interoperability between applicative, hardware and technological silos that may prevent from data exchange between different devices.

3.1.1. Objectives and methodology

To reach our overall goal, we will pursue the two following objectives, similar to the ones we set for the previous evaluation period. These two objectives are othogonal and can be carried on jointly :

1. Providing realistic complete self-organizing tools *e.g. vertical perspective*.
2. Going to heterogeneous energy-efficient performing wireless networks *e.g. horizontal perspective*,

We give more details on these two objectives below. To achieve our main objectives, we will mainly apply the methodology depicted in Figure 2 combining both theoretical analysis and experimental validation. Mathematical tools will allow us to properly dimension a problem, formally define its limitations and needs to provide suitable protocols in response. Then, they will allow us to qualify the outcome solutions before we validate and stress them in real scenarios with regards to applications requirements. For this, we will realize proofs-of-concept with real scenarios and real devices. Differences between results and expectations will be analyzed in return in order to well understand them and integrate them by design for a better protocol self-adaptation capability.

3.2. Vertical Perspective

As mentioned, future ubiquitous networks evolve in dynamic and unpredictable environments. Also, they can be used in a large scope of applications that have several expectations in terms of performance and different contextual limitations. In this heterogeneous context, IoT devices must support multiple applications and relay traffic with non-deterministic pattern.

To make our solutions practical and efficient in real conditions, we will adopt the dual approach both *top-down* and *bottom-up*. The *top-down* approach will ensure that we consider the application (such as throughput, delay, energy consumption, etc) and environmental limitations (such as deployment constraints, etc). The *bottom-up* approach will ensure that we take account of the physical and hardware characteristics such as memory, CPU, energy capacities but also physical interferences and obstacles. With this integrated perpective, we will be in capacity to design **cross-layer** integrated protocols well adapted [39]. We will design jointly routing and MAC layers by taking dynamics occurring at the physical layer into account with a constant concern for energy and security. We will investigate new adaptive frequency hopping techniques combined with routing protocols [41], [50], [24]. Also, we will work on new scheduling techniques for TSCH (a MAC layer of IEEE 802.15.4e) that are able work under the above-mentionned assumptions and bring the robustness of TSCH to IoT scenarios. We will investigate the performance boundaries of TSCH in particular in terms of energy-efficiency of time synchronization [63], and will propose alternative approaches such as capture effect-based time synchronization in TSCH or opportunistic routing. Another technology we will consider is IEEE 802.15.4g, which provides communication ranges in the order of tens of kilometers. We will propose

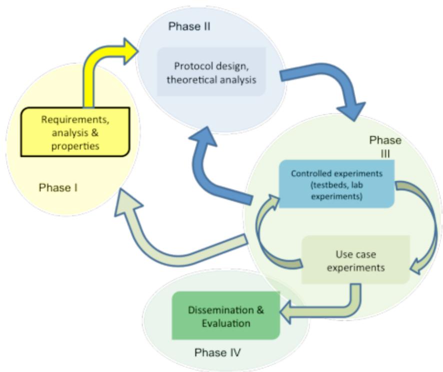


Figure 2. Methodology to be applied in FUN.

mechanisms to support scaling up to networks with a density of hundreds of nodes, at the MAC layer and above. We will also consider dual-technology networks where both long and short-range communication cooperate for increased robustness.

This vision will also allow us to integrate external factors by design in our protocols, in an opportunistic way. Yet, we will leverage on the occurrence of any of these phenomena rather than perceiving them as obstacles or limitations. As an example, we will rely on node undergone mobility to enhance routing performance as we have started to investigate in [74], [59]. On the same idea, when specific features are available like controlled mobility, we will exploit it to improve connectivity or coverage quality like in [46] [67].

3.3. Horizontal perspective

We aim at designing efficient tools for a plethora of wireless devices supporting highly heterogeneous technologies. We will thus investigate these networks from a horizontal perspective, e.g. by considering heterogeneity in low level communications layers.

Given the spectrum scarcity, they will probably need to coexist in the same frequency bands and sometimes for different purposes (RFID tag reading may use the same frequency bands as the wireless sensors). One important aspect to consider in this setting is how these different access technologies will interact with each other and what are the mechanisms needed to be put in place to guarantee that all services obtain the required share of resources when needed. This problem appears in different application domains, ranging from traffic offloading to unlicensed bands by cellular networks and the need to coexist with WiFi and radars, from a scenario in which multiple-purpose IoT clouds coexist in a city [75]. We will thus explore the dynamics of these interactions and devise ways to ensure smooth coexistence while considering the heterogeneity of the devices involved, the access mechanisms used as well as the requirements of the services provided.

To face the spectrum scarcity, we will also investigate new alternative communication paradigms such as phonon-based or light-based communications as we have initiated in [70], [71][16] and we will work on the coexistence of these technologies with traditional communication techniques, specifically by investigating efficient switching techniques from one communication technology to the other (they were most focused on the security aspects, to prevent jamming attacks). Resilience and reliability of the whole system will be the key factors to be taken into account [50], [24].

As a more prospective activity, we consider exploring software and communication security for IoT. This is challenging given that existing solutions do not address systems that are both constrained and networked [63]. Finally, in order to contribute to a better interoperability between all these technologies, we will continue to contribute to standardization bodies such as IETF and EPC Global.

4. Application Domains

4.1. Application Domains

The set of applications enabled through FUN and IoT is very large and can apply in every application area. We can thus not be exhaustive but among the most spread applications, we can name every area, event or animal monitoring, understanding and protection. To illustrate this, we may refer to the use cases addressed by our PREDNET project which goals is to equip rhinoceros with smart communicating devices to fight against poaching.

Other field of application is exploration of hostile and/or unknown environment by a fleet of self-organizing robots that cooperate with RFID and sensors to ensure a continue monitoring afterwards.

Also, IoT and FUN ca play a key role in logistics and traceability by relying on the use of sensors or RFID technologies as implemented in our TRACAVERRE project or our collaboration with the start up TRAXENS.

Finally, IoT and FUN leverage a lot of applications in Smart City concept , ranging from parking aid to a better energy consumption going through air quality monitoring, traffic fluidizing etc. (See our CityLab Inria and VITAL projects).

5. Highlights of the Year

5.1. Highlights of the Year

- The FIT facility has become an "Infrastructure de Recherche" (Infrastructure for Research) by the CD TGIR.

5.1.1. Awards

- Aziz Mbacke and Jad Nassar won the of the Hackaton at the SenZations summer school 2016, which opened them the doors of the UpRise Festival (http://uprisefestival.co/).
- Best paper award at the PIMRC 2016 conference .
- Viktor Toldov recipient of the award "Pepite 2016": (http://www.enseignementsup-recherche.gouv.fr/cid108805/3e-edition-du-prix-pepite-tremplin-pour-l-entrepreneuriat-etudiant-53-projets-recompenses.html)

BEST PAPERS AWARDS :

[47] **27th annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)**. V. TOLDOV, L. CLAVIER, V. LOSCRÍ, N. MITTON.

6. New Software and Platforms

6.1. FIT IoT-Lab

Participants: Nathalie Mitton [correspondant], Julien Vandaele.

FIT IoT-LAB is a very large scale open testbed that features over 2700 wireless sensor nodes and more than 200 robots spread across six different sites in France. Nodes are either fixed or mobile and can be allocated in various topologies throughout all sites. A variety of wireless sensors are available, with different processor architectures (MSP430, STM32 and Cortex-A8) and different wireless chips (802.15.4 PHY at 800 MHz or 2.4 GHz). In addition, "open nodes" can receive custom wireless sensors for inclusion in IoT-LAB testbed. This platform is completely open and can be used by any one wishing to run experiment on wireless sensors and robots.

The Lille site displays 3 subsets of the platforms:

- Euratechnologies : this site features 256 WSN430 sensor nodes operating in the 2.4GHz band. 64 nodes are mobile, embedded on mobile trains.
- Haute Borne : this site features 256 M3 sensor nodes operating in the 2.4GHz band and 64 mobile robots (32 turtlebots and 32 wifibots) completely remotely programmable.
- Opennodes : this site will feature (opening beginning 2015) 64 hardware open slots to allow any one to plug his own hardware and benefits from the platform debugging and monitoring tools.

7. New Results

7.1. Routing

Participants: Nathalie Mitton, Mouna Masmoudi.

Geographic routing is an attractive routing strategy in wireless sensor networks. It works well in dense networks, but it may suffer from the void problem. For this purpose, a recovery step is required to guarantee packet delivery. Face routing has widely been used as a recovery strategy since proved to guarantee delivery. However, it relies on a planar graph not always achievable in realistic wireless networks and may generate long paths. In [23], [12], we propose GRACO, a new geographic routing algorithm that combines a greedy forwarding and a recovery strategy based on swarm intelligence. During recovery, ant packets search for alternative paths and drop pheromone trails to guide next packets within the network. GRACO avoids holes and produces near optimal paths. Simulation results demonstrate that GRACO leads to a significant improvement of routing performance and scalability when compared to the literature algorithms.

GRACO has first been designed in the general case. We then studied its applicability to the Virtual Power Plants and their specific data packets with different priorities [23], [12]. Indeed, the Smart Grid (SG) incorporates communication networks to the conventional electricity system in order to intelligently integrate distributed energy resources (DERs) and allow for demand side management. The move to Smart grid in developing countries has to cope with great disparities of ICT infrastructures even within the same city. Besides, individual DERs are often too small to be allowed access to energy market, likewise power utilities are unable to effectively control and manage small DERs. We propose the use of affordable and scalable wireless communication technology to aggregate geographically sparse DERs into a single virtual power plant. The enrollment of prosumers in the VPP is conditional to financial performance of the plant. Thus, the VPPs are dynamic and are expected to scale up as more and more prosumers are attracted by their financial benefits. the communication network has to follow this progression and therefore to be scalable and rapidly deploy-able. We present a routing algorithm for data communication within the VPP to support centralized, decentralized or fully distributed control of the VPP's DERs.

Based on this study, we adapted GRACO so it can fit the specific cases of Smart Grid [23], [12] and more specifically to the Neighbor Area Networks (NAN) of Smart Grids, or distribution segment of the power system in the smart grid (SG). The deployment of ICT to support conventional grid will solve legacy problems that used to prevent implementation of smart services such as smart metering, demand side management or the integration of Distributed Energy Resources (DERs) within the smart grid. We demonstrate the effectiveness of GRACO in terms of scalability, peer-to-peer routing, end-to-end delay and delivery rate.

In another context, we made the observation that typical betweenness centrality metrics neglect the potential contribution of nodes that are near but not exactly on shortest paths. The idea of [35] is to give more value to these nodes. We propose a weighted betweenness centrality, a novel metric that assigns weights to nodes based on the stretch of the paths they intermediate against the shortest paths. We compare the proposed metric with the traditional and the distance-scaled betweenness metrics using four different network datasets. Results show that the weighted betweenness centrality pinpoints and promotes nodes that are underestimated by typical metrics, which can help to avoid network disconnections and better exploit multipath protocols.

7.2. Cloud and IoT

Participants: Valeria Loscri, Nathalie Mitton, Riccardo Petrolo.

Innovative and effective solutions to the fragmentation issues in the Internet of Things (IoT) landscape have been designed and proof of concept have been implemented to show the feasibility and effectiveness of the Cloud of Things (CoT) paradigm. In other words, we have focused on the convergence of Web semantic technologies and the Cloud computing concept as key enabler of an horizontal integration of various IoT applications and platforms [21]. The heterogeneity has to be considered not only in terms of applications and platforms, but another "type of heterogeneity" that deserves to be considered and analyzed is based on different devices and their interoperability.

A feasible solution to make different and heterogeneous devices to "interoperate" is based on the exploitation of a gateway. In particular, we have considered a Gateway-as-a-Service (Gaas) in [36], where we have shown that it is an efficient and lightweight device, which can be shared between several final users. Through the container virtualization technologies, we have been able to show how several platform requirements can be met, in a context where constrained devices have been considered. This study has demonstrated the Gateway-as-a-Service (GaaS) effectiveness and its exploitability in several IoT contexts, such as smart home, buildings, farms, agriculture environments, etc.

A different and complementary, to the previous solutions, perspective of IoT paradigm is represented by the management of the huge amount of data that have to be treated in the different IoT based applications. In [45], an infer algorithm has been proposed and more specifically an Bayesian Inference Approach (BIA) with the amin objective to avoid the transmission of high spatio-temporal correlated data.

7.3. Resource management in FUN

Participants: Cristina Cano Bastidas, Valeria Loscri, Simon Duquennoy.

A standard solution for reliable low-power mesh networks was defined in IEEE802.15.4e-2012, through the new MAC layer TSCH. TSCH (Time-Slotted Channel Hopping) provides a globally synchronized network that enables scheduling and channel hopping. Our review paper [28] details the TSCH technology as well as the 6LoWPAN and 6TiSCH protocols. It gathers authors from all major open-source IoT OSes: Contiki, OpenWSN, RIOT and TinyOS. The paper presents architectural considerations when it comes to implementing portable TSCH stacks, and presents preliminary evaluation results.

TSCH networks require global synchronization. The more precise the synchronization, the more energy-efficient the network. We address the challenge of reaching micro-second time synchronization over multiple hops in TSCH networks [31], at low power. The key idea is to use two crystal oscillators, one at low-frequency for low-power timekeeping, one at high-frequency for intra-slot precision. Along with adaptive drift compensation, this method is proven effective through an experimental assessment.

Beaconing is usually employed to allow network discovery and to maintain synchronisation in mesh networking protocols, such as those defined in the IEEE 802.15.4e and IEEE 802.11s standards. Thus, avoiding persistent or consecutive collisions of beacons is crucial in order to ensure correct network operation. Beacons are also used in receiver-initiated medium access protocols to advertise that nodes are awake. Consequently, effective beacon scheduling can enable duty-cycle operation and reduce energy consumption. We propose [56] a completely decentralised and low-complexity solution based on learning techniques to schedule beacon transmissions in mesh networks. We show the algorithm converges to beacon collision-free operation almost surely in finite time and evaluate converge times in different mesh network scenarios.

In [54] we focus on new methods, architectures, and applications for the management of Cyber Physical Objects (CPOs) in the context of the Internet of Things (IoT). The book covers a wide range of topics related to CPOs, such as resource management, hardware platforms, communication and control, and control and estimation over networks. It also discusses decentralized, distributed, and cooperative optimization as well as effective discovery, management, and querying of CPOs. Other chapters outline the applications of control, real-time aspects, and software for CPOs and introduce readers to agent-oriented CPOs, communication support for CPOs, real-world deployment of CPOs, and CPOs in Complex Systems. There is a focus on the importance of application of IoT technologies for Smart Cities.

Finally, we address software security and in particular the challenge of formally verifying the source code of IoT OSes. This is the topic of the yet-to-be-started H2020 VESSEDIA project. Our preliminary study [32] demonstrated the feasibility of applying Frama-C to a memory allocation module of the Contiki OS.

7.4. Smart Cities

Participants: Nathalie Mitton, Valeria Loscri, Riccardo Petrolo.

Smart City represents one of the most promising, prominent and challenging Internet of Things (IoT) applications, but recent ICT trends suggest more and more that cities could also benefit from Cloud computing. The convergence of IoT paradigm and Cloud computing technology, can play a fundamental role for developing of highly level and organized cities form an ICT point of view, but it is of paramount importance to deal a critical analysis to identify the issues and challenges deriving from this synergy.

A novel perspective that we have considered as key factor for the realization of Future Internet is the role of the interconnected objects as active entities in the context of the networked systems [52]. With this perspective in mind, we have proposed CACHACA [43], a ranking mechanism for Sensor Networks that facilitate the discovery of services provided by each network element. Discovery functionality has been also considered in the context of VITAL project, since effective and accurate mechanisms to discover Inter-Connected Objects (ICOs) and new services represents a sine qua non condition to have effective exploration of data-sources that are appropriate for a specific business context as defined by an end-user [42] [11].

On the other hand, a Smart City is a kind of ecosystem characterized with different IoT solutions that have to cooperate and coexist and is in continuous expansion. In order to face with the integration and interoperability challenges of this ecosystem, we have considered VITAL-OS architecture that can monitor, visualize, and control all the operations of a city [44].

7.5. RFID

Participants: Nathalie Mitton, Abdoul Aziz Mbacke.

One of the devices under consideration by the FUN team is RFID. One of the main issues to widely deploy RFID reader is reader-to-reader collision. Indeed, when the electromagnetic fields of the readers overlap, a collision occurs on the tag laying in the overlapping section and cannot be read. Numerous protocols have been proposed to attempt to reduce them, but, remaining reading errors still heavily impact the performances and fairness of dense RFID deployments. In [33], [18] we introduce a new Distributed Efficient & Fair Anticollision for RFID (DEFAR) protocol. It reduces both monochannel and multichannel collisions as well as interference by a factor of almost 90% in comparison with the best state of the art protocols. The fairness of the medium access among the readers is improved to a 99% level. Such improvements are achieved applying a TDMA-based "server-less" approach and assigning different priorities to readers depending on their behavior over precedent rounds. A distributed reservation phase is organized between readers with at least one winning reader afterwards. Then, multiple reading phases occur within a single frame in order to obtain fast coverage and high throughput. The use of different reader priorities based on reading behaviors of previous frames also contributes to improve both fairness and efficiency. Simulation results show the robustness of the proposed solution in terms of different metrics such collision avoidance, fairness and coverage and in comparison with a centralized literature solution.

In order to ensure collision-free reading, a scheduling scheme is needed to read tags in the shortest possible time. We study in [37] this scheduling problem in a stationary setting and the reader minimization problem in a mobile setting. We show that the optimal schedule construction problem is NP-complete and provide an approximation algorithm that we evaluate our techniques through simulation.

7.6. Interferences and failures management

Participants: Nathalie Mitton, Viktor Toldov, Valeria Loscri, Simon Duquennoy.

In the recent years, the Machine-to-Machine (M2M) paradigm together with the integration of wireless sensors networks with the generic infrastructure via $6LoWPAN$ require the implementation of ad hoc communication protocols at the Medium Access Control layer, that do not depend on pre-existing infrastructure. Channel hopping concept has more and more gained consensus as a viable and effective solution for wireless MAC layer coordination with time-synchronized channel hopping (TSCH). In [24] we propose a decentralized multichannel MAC coordination framework (DT-SCS) leveraging the concept of *pulse-coupled oscillators* at the MAC layer. In DT-SCS, nodes randomly join a channel and are automatically spread across the available channels. The nodes then achieve PCO-based coordination via the periodic transmission of beacon packets at the MAC layer. As such, for channels with an equal number of nodes, DT-SCS converges to synchronized beacon packet transmission at the MAC layer in a completely uncoordinated manner. In order to combat the well-know phenomenon of Cross-Technology Interference (CTI) a cross-layer mechanism, CrossZig, has been implemented in [39], based on the exploitation of information at the physical layer in order to detect the presence of CTI in a corrupted packet.

A different perspective of the interference management has been considered in [47] and [41], where a novel solution to allow to secondary users the access of allocated spectrum has been proposed. The study has been based on the major consideration that a big bottleneck in cognitive radio systems is based on finding the best available channel as fast as possible.

A totally different approach to face the enormous quantity of data generated by IoT devices, is to try to reduce the sending of useless data, based on the adoption of effective predictive approaches.

In [50] we have considered the concept of high spatio-temporal correlated data and we have proposed a Belief Propagation (BP) algorithm to derive methods to drastically reduce the number of transmitted messages, by keeping an high accuracy in terms of global information.

Together with interference management approaches it is also important to figure out tools to support network operator for mitigation of the impact of failures on their infrastructures. The need of advanced Network Planning and Management Tool (NPMT) has been considered in [30].

7.7. Vehicular Networks

Participants: Nathalie Mitton, Valeria Loscri.

[27] studies the information delivery delay analysis for roadside unit deployment in a vehicular ad hoc network (VANET) with intermittent connectivity. A mathematical model is developed to describe the relationship between the average delay for delivering road condition information and the distance between two neighbor RSUs deployed along a road. The derived mathematical model considers a straight highway scenario where two RSUs are deployed at a distance without any direct connection and vehicles are sparsely distributed on the road with road condition information randomly generated between the two neighbor RSUs. Moreover, the model takes into account the vehicle speed, the vehicle density, the likelihood of an incident, and the distance between two RSUs. The effectiveness of the derived mathematical model is verified through simulation results. Given the information delivery delay constraint of a time-critical application, this model can be used to estimate the maximum distance allowed between two neighbor RSUs, which can provide a reference for the deployment of RSUs in such scenarios.

But Vehicular Networks can also convey social networks. In [53], we survey recent literature on Vehicular Social Networks that are a particular class of vehicular ad hoc networks, characterized by social aspects and features. Starting from this pillar, we investigate perspectives of next generation vehicles under the assumption of social networking for vehicular applications (i.e., safety and entertainment applications). This paper plays a role as a starting point about socially-inspired vehicles, and main related applications, as well as communication techniques. Vehicular communications can be considered as the "first social network for automobiles", since each driver can share data with other neighbors. As an instance, heavy traffic is a common occurrence in some areas on the roads (e.g., at intersections, taxi loading/unloading areas, and so on); as a consequence, roads become a popular social place for vehicles to connect to each other. Human factors are then involved in vehicular ad hoc networks, not only due to the safety related applications, but also for entertainment purpose. Social characteristics and human behavior largely impact on vehicular ad hoc networks, and this arises to the vehicular social networks, which are formed when vehicles (individuals) "socialize" and share common interests. This survey describes the main features of vehicular social networks, from novel emerging technologies to social aspects used for mobile applications, as well as main issues and challenges. Vehicular social networks are described as decentralized opportunistic communication networks formed among vehicles. They exploit mobility aspects, and basics of traditional social networks, in order to create novel approaches of message exchange through the detection of dynamic social structures. An overview of the main state-of-the-art on safety and entertainment applications relying on social networking solutions is also provided.

Cognitive Radio (CR) together with vehicular networks have been considered with an integrated and synergic perspective in [55], since CR technology is foreseen as a very effective tool to improve the communication efficiency in the context of vehicular networked systems.

7.8. Self-deployment and coverage

Participants: Nathalie Mitton, Tahiry Razafindralambo.

Controlled mobility in wireless sensor networks can provide many services. One of the most challenging one is coverage. Coverage can be needed either for monitoring control of specific area or point of interest or for deploying a communication network. This latter case is required for instance in post-disaster situations. In post-disaster scenarios, for example, after earthquakes or floods, the traditional communication infrastructure may be unavailable or seriously disrupted and overloaded. Therefore, rapidly deployable network solutions are needed to restore connectivity and provide assistance to users and first responders in the incident area. This work surveys the solutions proposed to address the deployment of a network without any a priori knowledge about the communication environment for critical communications. The design of such a network should also allow for quick, flexible, scalable, and resilient deployment with minimal human intervention. We survey this kind of approaches in [20].

In [13], we present a decentralized deployment algorithm for wireless mobile sensor networks focused on deployment Efficiency, connectivity Maintenance and network Reparation (EMR). We assume that a group of mobile sensors is placed in the area of interest to be covered, without any prior knowledge of the environment. The goal of the algorithm is to maximize the covered area and cope with sudden sensor failures. By relying on the locally available information regarding the environment and neighborhood, and without the need for any kind of synchronization in the network, each sensor iteratively chooses the next-step movement location so as to form a hexagonal lattice grid. Relying on the graph of wireless mobile sensors, we are able to provide the properties regarding the quality of coverage, the connectivity of the graph and the termination of the algorithm. We run extensive simulations to provide compactness properties of the deployment and evaluate the robustness against sensor failures. We show through the analysis and the simulations that EMR algorithm is robust to node failures and can restore the lattice grid. We also show that even after a failure, EMR algorithm call still provide a compact deployment in a reasonable time.

Routing a fleet of robots in a known surface is a complex problem. It consists in the determination of the exact trajectory each robot has to follow to collect information. The objective pursued in [38] is to maximize the exploration of the given surface. To ensure the robots can execute the mission in a collaborative manner, connectivity constraints are considered. These constraints guarantee that robots can communicate among each other and share the collected information. Moreover, the trajectories of the robots need to respect autonomy constraints.

7.9. Controlled Mobility for additional services

Participants: Nathalie Mitton, Valeria Loscri, Jean Cristanel Razafimandimby Anjalalaina.

Wireless sensor networks (WSNs) have been of very high interest for the research community since years, but most of the time, the mobility of nodes have been considered as an obstacle to overcome. In the contrary, in have tried to adopt another perspective and see it as an asset to exploit to provide additional services.

In [19], we leverage on the ability of mobile nodes to replace or recharge static sensors. Two main approaches can be identified that target this objective: either “recharging” or “replacing” the sensor nodes that are running out of energy. Of particular interest are solutions where mobile robots are used to execute the above mentioned tasks to automatically and autonomously maintain the WSN, thus reducing human intervention. Recently, the progress in wireless power transfer techniques has boosted research activities in the direction of battery recharging, with high expectations for its application to WSNs. Similarly, also sensor replacement techniques have been widely studied as a means to provide service continuity in the network. Objective of [19] is to investigate the limitations and the advantages of these two research directions. Key decision points must be identified for effectively supporting WSN self-maintenance: (i) which sensor nodes have to be recharged/replaced; (ii) in which order the mobile robot is serving (i.e., recharging/replacing) the nodes and by following which path; (iii) how much energy is delivered to a sensor when recharged. The influence that a set of parameters, relative to both the sensors and the mobile robot, on the decisions will be considered. Centralized and distributed solutions are compared in terms of effectiveness in prolonging the network lifetime and in allowing network self-sustainability. The performance evaluation in a variety of scenarios and network settings offers the opportunity to draw conclusions and to discuss the boundaries for one technique being preferable to the other.

Mobility can also help for collecting data in wireless sensor networks [29]. The sensor data collection problem using data mules have been studied fairly extensively in the literature. However, in most of these studies, while the mule is mobile, all sensors are stationary. The objective of most of these studies is to minimize the time needed by the mule to collect data from all the sensors and return to the data collection point, from where it embarked on its data collection journey. The problem studied in this paper has two major differences with the earlier studies. First, in this study we assume that both the mule as well as the sensors are mobile. Second, we do not attempt to minimize the data collection time. Instead we minimize the number of mules that will be needed to collect data from all the sensors, subject to the constraint that the data collection process has to be completed within some pre-specified time. We show that the mule minimization problem is NP-Complete and provide a solution by first transforming it to a generalized version of the minimum flow problem in a network and then solving it optimally using Integer Linear Programming. Finally, we evaluate our algorithms through extensive simulation and present the results.

Internet of Robotic Things (IoRT) is a new concept introduced for the first time by ABI Research. Unlike the Internet of Things (IoT), IoRT provides an active sensorization and is considered as the new evolution of IoT.

This new concept will bring new opportunities and challenges, while providing new business ideas for IoT and robotics' entrepreneurs.

In [46], we focus particularly on two issues: (i) connectivity maintenance among multiple IoRT robots, and (ii) their collective coverage.

We propose (i) IoRT-based, and (ii) a neural network control scheme to efficiently maintain the global connectivity among multiple mobile robots to a desired quality-of-service (QoS) level. The proposed approaches will try to find a trade-off between collective coverage and communication quality.

The IoT-based approach is based on the computation of the algebraic connectivity and the use of virtual force algorithm.

The neural network controller, in turn, is completely distributed and mimics perfectly the IoT-based approach. Results show that our approaches are efficient, in terms of convergence, connectivity, and energy consumption.

7.10. New and other communication paradigms

Participants: Nathalie Mitton, Valeria Loscri.

Interconnection and self-organized systems are normally populated with heterogeneous and different devices. The differences range from computational capabilities, storage size, etc. Instead of considering the heterogeneity as a limitation, it is possible to "turn it" as a primitive control of the system, in order to realize more robust and more resilient communication systems.

Based on those considerations, we have studied and analyzed the specific features of devices belonging to the category of micro-nano nodes that are however, required to interact with up-sized devices.

In order to improve the understanding of the behavior of micro/nano-sized devices, we have considered fundamental the analysis in specific applications and environment, where this kind of devices can be largely exploited, such as on/in-body networks applications.

Indeed, we retain that bio-medical applications can be advantaged by an effective and efficient communication and cooperation of devices deployed both on top of the body and inside it. Even if the research community recognizes a great importance to the study of interaction between the Human Immune System (HIS) and nano devices, this branch of research is in its infancy due to the major issue to model the HIS. A theoretical derivation of HIS and its interaction with a nanoparticulate system have been proposed in [15]. Some experimental results have been derived in [16], where specific parameters, e.g. temperature variations, Ph, etc. have been considered to establish the biocpmpatibility of TiO2 particles with human tissues.

A step ahead in this direction has consisted in the consideration of alternative particles as potential information carriers always in the context of biological environments. In [40] we have studied *phonons* as information carriers, we have derived a channel modeling and evaluated the theoretical capacity. The main reasons for taking into consideration this type of nanoparticles are twofold. Firstly, phonons represent something that is naturally generated in a biological context with the application of a tolerable electromagnetic field and secondly they represent a straightforward way to implement nanomachines, since their native size.

7.11. Modelling and experimentations of interferences and other PHY effects

Participants: Nathalie Mitton, Valeria Loscri.

In the era of Internet of Things (IoT), the development of Wireless Sensor Networks (WSN) arises different challenges. Two of the main issues are electromagnetic interference and the lifetime of WSN nodes. In [48], we show and evaluate experimentally the relation between interference and energy consumption, which impacts the network lifetime. We present a platform based on commercially available low-cost hardware in order to evaluate the impact of electromagnetic interference in 2.4 GHz ISM band on energy consumption of WSN. The energy measurements are obtained separately from each electronic component in the node. Interference and energy measurements are conducted in an anechoic chamber and in an office-type lab environment. X-MAC protocol is chosen to manage the Radio Duty Cycle of the nodes and its energy performance is evaluated. The energy consumption transmitter nodes is analyzed particularly in this work. Moreover, this energy consumption has been quantified and differentiated according to the number of (re-)transmissions carried out by the transmitter as well as the number of ACK packets sent by the receiver for a single packet. Finally, we use a model of real battery to calculate the lifetime of the node for operation within different interference level zones. This study lays the basis for further design rules of communication protocols and development of WSNs.

In [49], we propose a WSN architecture for wild animal monitoring. The key requirements of the system are long range transmissions and low power consumption. Indeed, the animals could be spread over vast areas. Kruger National Park in South Africa (19485 km2) is the potential zone of implementation of the network. On the other hand, size and weight limitations of wearable devices must be respected, which limits the size and capacity of battery. Moreover, battery replacement is a difficult and expensive process. So, low energy consumption is essential to extend the network lifetime. Some animal tracking projects [3] use GSM to transmit collected data to insure the coverage over a large area. However, high energy consumption of GSM and lack of coverage of the deployment area do not meet the essential requirements of the application. LoRa technology provides both long range transmissions and low power operation. This technology could be an appropriate solution for PREDNET project. The contribution of this work is multiple: 1) we defined communication parameters of LoRa radio for PREDNET WSN; 2) we performed radio propagation simulation for chosen parameters to estimate the coverage area for both urban and wilderness (rural) scenarios; 3) we confirmed the propagation simulations with range tests; 4) we measured experimentally the Packet Error Rate (PER) of transmissions.

Terahertz frequency band is an emerging research area related to nano-scale communications. In this frequency range, specific features can provide the possibility to overcome the issues related to the spectrum scarcity and capacity limitation.

Apart high molecular absorption, and very high reflection loss that represent main phenomena in THz band, we can derive the characteristics of the channel affected by chirality effects occurring in the propagation medium, specifically , in the case where a Giant Optical Activity is present. This effect is typical of the so-called chiral-metamaterials in (4-10) THz band, and is of stimulating interest particularly for millimeter wireless communications.

In [51], [25], we analyze the behavior of specific parameters of a chiral-metamaterial, like the relative electrical permitivity, magnetic permeability and chirality coefficients, and from that we derive the channel behavior both for Line-of-Sight and No Line-of-Sight propagations. We notice the presence of spectral windows, due to peaks of resonance of chiral parameter.

Finally, performances of the chirality-affected channel have been assessed in terms of (i) channel capacity, (ii) propagation delay, and (iii) coherence band-width, for different distances.

Thanks to the exploitation of frequencies in the interval ranging from 0.06 to 10 THz, it is envisioned the possibility to overcome the issues related to the spectrum scarcity and capacity limitation. On the other hand, the design of new channel models, able to capture the inherent features of the phenomenons related with this specific field is of paramount importance. Very high molecular absorption, and very high reflection loss are peculiarities phenomenons that need to be included in these models. In [26], we present a full-wave propagation model of the electromagnetic field that propagates in the THz band both for Line-of-Sight and Non-Line-of-Sight propagation models. In the full-wave model, we also introduce the chirality effects occurring in the propagation medium, i.e., a chiral metamaterial.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Evolution
  Participants: Gabriele Sabatino, Nathalie Mitton [correspondant].

  This collaboration aims to set up a full RFID system on the basis of AspireRFID middleware and pre-existing RFID modules issued from FUN research in the Evolution company facility and to integrate them with their IS.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Tracaverre

Participants: Nathalie Mitton [correspondant], Gabriele Sabatino.

Title: Tracaverre

Type: FUI

Duration: November 2012 - Dec 2016

Coordinator: Saver Glass

Others partners: Inria FUN IEMN Courbon Camus La Grande Marque LIRIS DISP

Tracaverre studies the use of RFID for traceability of prestigious bottles. Tracaverre has yielded to the implementation of the T-Scan software.

9.1.2. StoreConnect

Participants: Nathalie Mitton [correspondant], Valeria Loscri.

Title: StoreConnect

Type: FUI

Duration: September 2016 - October 2018

Coordinator: NEOSENSYS

Others partners: Inria FUN, SPIRALS and STARS, TeVolys, Ubudu, Smile, STIME, Leroy Merlin

The aim of StoreConnect is to provide French large retailers with efficient and powerful tools in the in-store customer interaction.

9.1.3. PIPA

Participants: Nathalie Mitton [correspondant], Farouk Mezghani, Cristina Cano Bastidas.

Title: Partager de l'Info PArtout à bas coût

Type: Chercheur citoyen

Duration: Dec 2015 - Dec 2017

Coordinator: Inria FUN

Others partners: SpotTrotter

PIPA project aims to provide an innovative low cost solution to share information in places where communication infrastructure are lacking, insuffisant or not adapted, going beyond technical, economical or political limitations.

9.2. National Initiatives

9.2.1. Inria Project Lab

9.2.1.1. CityLab@Inria

Participants: Valeria Loscri, Abdoul Aziz Mbacke, Nathalie Mitton [correspondant].

- Title: CityLab@Inria
- Type: IPL
- Duration: 2015 - 2019
- Coordinator: Valerie Issarny
- CityLab@Inria studies ICT solutions toward smart cities that promote both social and environmental sustainability. A strong emphasis of the Lab is on the undertaking of a multi-disciplinary research program through the integration of relevant scientific and technology studies, from sensing up to analytics and advanced applications, so as to actually enact the foreseen smart city Systems of Systems. Obviously, running urban-scale experiments is a central concern of the Lab, so that we are able to confront proposed approaches to actual settings. The Lab's research leverages relevant effort within Inria project-teams that is further revisited as well as integrated to meet the challenges of smart cities. Research themes span: energy-efficient wireless communication protocols, urban-scale social and physical sensing, privacy by design, cloud-based urban data management, data assimilation, visual analysis, and urban system software engineering. In addition, CityLab Inria research builds upon collaborative effort at the International level, and especially collaboration in the context of the Inria SiliconValley program. This project has yield to the set up of a full course on Smart Cities via a MOOC.

9.2.2. ADT

9.2.2.1. RFunID

Participants: Clement Fumey, Nathalie Mitton [correspondant], Julien Vandaele.

Duration: September 2015 - August 2017

Coordinator: Inria FUN

The purpose of this project is to deploy a large scale experimental RFID platform that enables remote programmation of RFID scenario on heterogeneous devices.

9.2.2.2. ARUNTA

Participants: Emilio Compagnone, Valeria Loscri [correspondant], Julien Vandaele.

Title: Arduino-based Robots for Ubiquitous Network (ARUNTA)

Type: ADT

Duration: September 2014 - August 2016

Coordinator: Inria FUN

Abstract: This ADT focuses on the use of Arduino, an open-source electronics prototyping platform, really flexible and easy-to-use [1] to allow a fleet of robots to perform specific tasks. The goal of the ADT is to make experiments on Arduino-based robotic platforms, by implementing two robot cooperation algorithms that have been already tested through simulation tools. In order to extend the users' community and to allow more people to benefit from this research on robot cooperation, this ADT will output a tutorial and a test-bed will be developed. Moreover, the final project will be shared with the Arduino community and every interested user.

9.2.3. Equipements d'Excellence

9.2.3.1. FIT

Participants: Nathalie Mitton [correspondant], Julien Vandaele.

Title: Future Internet of Things

Type: EquipEx

Duration: March 2010 - December 2019

Coordinator: UPMC

See also: http://fit-equipex.fr/

Abstract: FIT (Future Internet of Things) aims to develop an experimental facility, a federated and competitive infrastructure with international visibility and a broad panel of customers. It will provide this facility with a set of complementary components that enable experimentation on innovative services for academic and industrial users. The project will give French Internet stakeholders a means to experiment on mobile wireless communications at the network and application layers thereby accelerating the design of advanced networking technologies for the Future Internet. FIT is one of 52 winning projects from the first wave of the French Ministry of Higher Education and Research's "Equipements d'Excellence" (Equipex) research grant program. Coordinated by Professor Serge Fdida of UPMC Sorbonne Universités and running over a nine-year period, the project will benefit from a 5.8 million euro grant from the French government. This project has yield to several publications in 2016: [44], [43].

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

9.3.1.1. VITAL

Participants: Salvatore Guzzo Bonifacio, Valeria Loscri, Nathalie Mitton [correspondant], Riccardo Petrolo.

Title: Virtualized programmable InTerfAces for innovative cost-effective IoT depLoyments in smart cities

Programm: FP7

Duration: September 2013 - December 2016

Coordinator: National University of Ireland Galway (NUIG)

Partners:

Research and Education Laboratory in Information Technologies (Greece)

Atos Spain (Spain)

Camden Town Center (United Kingdom)

Images & Co (United Kingdom)

Istanbul Metropolitan Municipality (Turkey)

Istanbul Teknik Universitesi (Turkey)

National University of Ireland, Galway (Ireland)

Santer Reply Spa (Italy)

Singularlogic Anonymi Etairia Pliroforiakon Sistimaton Kai Efarmogon Pliroforikis (Greece)

Inria contact: Nathalie Mitton

Internet-of-Things (IoT) applications are currently based on multiple architectures, standards and platforms, which have led to a highly fragmented IoT landscape. This fragmentation is evident in the area of smart cities, which typically comprise several technological silos (i.e. IoT systems that have been developed and deployed independently). Nowadays there is a pressing need to remove these silos in order to allow cities to share data across systems and coordinate processes across domains, thereby essentially improving sustainability and quality of life. In response to this need, VITAL will realize a radical shift in the development, deployment and operation of IoT applications, through introducing an abstract virtualized digital layer that will operate across multiple IoT architectures, platforms and business contexts. Specifically, VITAL will provide platform and business context agnostic access to Internet-Connected-Objects (ICO). Moreover, it will research virtualized filtering, complex event processing (CEP) and business process management mechanisms, which will be operational over a variety of IoT architectures/ecosystems. The mechanisms will compromise the diverse characteristics of the underlying ecosystems, thereby boosting interoperability at the technical and business levels. VITAL will also provide development and governance tools, which will leverage the project's interfaces for virtualized access to ICOs.VITAL will allow solution providers to (re)use a wider range of data steams, thereby increasing the scope of potential applications. It will also enable a more connected/integrated approach to smart city applications development, which will be validated in realistic deployments in London and Istanbul. The partners will contribute and adapt a host of readily available urban infrastructures, IoT platforms and novel IoT applications, which will ease the accomplishment of the project's goals based on an optimal value for EC money. Publications in 2016 in the framework of this project are: [21], [44], [43], [42].

9.3.1.2. VESSEDIA

Participant: Simon Duquennoy [correspondant].

Title: VERIFICATION ENGINEERING OF SAFETY AND SECURITY CRITICAL DYNAMIC INDUSTRIAL APPLICATIONS

Programm: H2020

Duration: January 2017 - Dec. 2019

TECHNIKON FORSCHUNGS UND PLANUNGSGESELLSCHAFT MBH (TEC) The VESSEDIA project will bring safety and security to many new software applications and devices. In the fast evolving world we live in, the Internet has brought many benefits to individuals, organisations and industries. With the capabilities offered now (such as IPv6) to connect billions of devices and therefore humans together, the Internet brings new threats to the software developers and VESSEDIA will allow connected applications to be safe and secure. VESSEDIA proposes to enhance and scale up modern software analysis tools, namely the mostly open-source Frama-C Analysis platform, to allow developers to benefit rapidly from them when developing connected applications. At the forefront of connected applications is the IoT, whose growth is exponential and whose security risks are real (for instance in hacked smart phones). VESSEDIA will take this domain as a target for demonstrating the benefits of using our tools on connected applications. VESSEDIA will tackle this challenge by 1) developing a methodology that allows to adopt and use source code analysis tools efficiently and produce similar benefits than already achieved for highly-critical applications (i.e. an exhaustive analysis and extraction of faults), 2) enhancing the Frama-C toolbox to enable efficient and fast implementation, 3) demonstrating the new toolbox capabilities on typical IoT (Internet of Things) applications including an IoT Operating System (Contiki), 4) developing a standardisation plan for generalizing the use of the toolbox, 5) contributing to the Common Criteria certification process, and 6) defining a label "Verified in Europe" for validating software products with European technologies such as Frama-C.

9.4. International Initiatives

9.4.1. Inria International Labs

9.4.1.1. PREDNET

Participants: Simon Duquennoy, Nathalie Mitton [correspondant], Viktor Toldov, Julien Vandaele.

Title: Predator network

Type: LIRIMA with Stellenbosch University, South Africa

Duration: January 2013 - December 2016

See also: https://iww.inria.fr/prednet/en/

Abstract: PREDNET (PREDator adhoc NETwork) proposes to do research on the most suitable topology and subsequent deployment of a wireless sensor network for sparsely populated outlying rural and wilderness areas, for effective monitoring and protection of resources and ecosystems. This collaboration gave birth to joint project submission, joint conference organization and several publications, among them for 2016 [47], [48], [48], [49].

9.4.2. Inria Associate Teams Not Involved in an Inria International Labs

9.4.2.1. DepIoT

Participants: Simon Duquennoy [correspondant], Nathalie Mitton.

Title: DepIoT: Coexistence and Security for Dependable Internet of Things

Type: North-European Inria Associate Team with SICS, Sweden

Duration: Sept 2016 - August 2018

Abstract: In order to foster the adoption of IoT technologies, dependability must be guaranteed. We will tackle this challenge by ensuring operation even in the presence of other networks sharing the same frequency band (coexistence) and by enabling a secure communication.

9.4.3. Inria International Partners

9.4.3.1. Declared Inria International Partners

Université Mediterranea di Reggio Calabria (UNIC) (Italy) Objective of this collaboration is the design of an innovative architecture that enables autonomic and decentralized fruition of the services offered by the network of smart objects in many heterogeneous and dynamic environments, in a way that is independent of the network topology, reliable and flexible. The result is an 'ecosystem' of objects, self-organized and self-sustained, capable of making data and services available to the users wherever and whenever required, thus supporting the fruition of an 'augmented' reality thanks to a new environmental and social awareness. This collaboration has allowed students and researchers exchanges and joint publications, among them for 2016: [20], [19].

9.4.3.2. Informal International Partners

Southern University, China

The purpose of this collaboration is to study the green (or energy-efficient) communication problem in vehicular ad hoc networks (VANETs) and the application of vehicular network communication in green transportation. In this framework, Nathalie Mitton visited the Nanjing University. It gave birth to joint project submission, joint conference organization (UIC 2016) and joint publications, one in 2016 [27].

Arun Sen from Arizona State University, USA

The purpose of this collaboration is to study the joint scheduling and trajectory of RFID readers in a mobile environment. In this framework, Arun Sen visited the FUN team for 6 months in 2015 and in July 2016. It gave birth to joint project submission, joint conference submission and joint publications, among them in 2016 [30], [29].

Anna-Maria Vegni from Roma Tre University, Italy

The purpose of this collaboration is to study alternative communication paradigms and investigate their limitations and different effects on performances. In this framework, joint publications have been obtained, among them in 2016 [51], [26], [53], [50], [46], [40], [16], [55], [15], [45], [25].

9.4.3.3. PhD co-supervision

Participants: Nathalie Mitton [correspondant], Mouna Masmoudi.

Since January 2013, Nathalie Mitton co-supervises Mouna Rekik as a PhD student with Pr Zied Chtourou from Université de Sfax, Tunisia. Her topic is about swarm intelligence based multi-path geographic routing for wireless sensor and actuator networks.

This work has led to the following publications in 2016: [23]. Mouna defended her PhD on July 26th 2016.

Since 2014, Simon Duquennoy co-supervised Anwar Hithnawi as a PhD student with Pr Friedemann Mattern from ETH Zurich, Sweden. Her research is on low-power wireless systems coexistence, and mitigation of cross-technology inteference. This work has led to the following publications in 2016: [39]. Anwar defended her PhD on November 8, 2016.

9.4.4. Participation in Other International Programs

9.4.4.1. CROMO

Participants: Valeria Loscri, Nathalie Mitton [correspondant], Riccardo Petrolo, Abdoul Aziz Mbacke.

Title: Crowd Data In the mobile cloud

Duration: January 2015 - December 2019

CroMo (Crowd Data In the mobile cloud) is a submission to the CAPES-COFECUB project call lead by Inria from the French side and University of Rio de Janeiro from Brazilian Side. Other partner institutions are Université Pierre et Marie Curie and Université de la Rochelle.

Mobile cloud computing is an emerging paradigm to improve the quality of mobile applications by transferring part of the computational tasks to the resource-rich cloud. The multitude data sources combined with the known difficulties of wireless communications represent an important issue for mobile cloud computing. Therefore, the additional computational power added by the cloud has to deal with the constraints of the wireless medium. One could imagine a situation where different sensors collect data and require intensive computation. This data must be transmitted at high rates before becoming stale. In this case, the network becomes the main bottleneck, not the processing power or storage size. To circumvent this issue, different strategies can be envisioned. As usual alternatives, wireless data rates must be increased or the amount of data sent to the cloud must be reduced. CROMO tackles challenges from all these three components of the mobile clouds (data generation, collect and processing) to then integrate them as a whole enhanced mobile cloud with improved network performances in terms of delay, energy consumption, availability, and reliability.

In this context, joint exchanges and crossed visits have been done (Aziz went to Rio, Dianne went to Lille). The project yield to several publications such as [35], [36], [37].

9.5. International Research Visitors

9.5.1. Visits of International Scientists

9.5.1.1. Senior researchers

Several researchers have visited our group in 2016, mainly from our partner universities but not only:

- Zied Chtourou, Univ. Sfax, Tunisia, July 2016
- Arun Sen, Arizona State University, USA, July 2016
- Ahmet Sekercioglu, Monash University, Australia, July 2016
- Riaan Wolhuter, Univ. Stellenbosch, South Africa, July 2016
- Anwar Hithnawi, ETH Zurich, Switzerland, March 2016
- Hossein Shafagh, ETH Zurich, Switzerland, March 2016
- Cédric Chauvenet from ERDF, May 2016

9.5.1.2. Internships

Other students have visited us from our partner universities in the framework of the joint project we run together. This is the case for William Pretorius (2 months) who came from Stellenbosch university, South Africa, in the framework of the PredNET program and Rahul Vyas from IIIT Allahabad, India (2 months).

9.5.2. Visits to International Teams

- Riccardo Petrolo visited Ericsson group in Finland in April 2016.
- Nathalie Mitton visited Southeast university in Nanjing, China in June 2016.
- Simon Duquennoy visited SICS in Sweden in September 2016 and May 2016.
- Aziz Mbacke visited UFRJ in Brazil in November 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Nathalie Mitton is a member of the Steering committee of CIoT.
- Nathalie Mitton is general chair for AdHocNow 2016 and InterIoT 2016 and symposium chair of Wireless Ad hoc and Sensor Networks symposium ICNC 2017.
- Valeria Loscri' is TPC chair for BODYNETS 2016, general co-chairs for PASC 2016, TPC co-chair for CoWPER 2016, TPC chair for the IoT day 2016
- Mouna Rekik was publicity chair for AdHocNet 2016.
- Aziz Mbacke is publicity chair for AdHocNow 2016.
- Riccardo Petrolo was web chair for AdHocNow 2016.
- Jean Razafimandimby Anjalalaina, was Publicity Chair for IoTIP 2016.

10.1.1.2. Member of the Organizing Committees

- the next CapTronic workshop on Factories of the future. Attendees : industries, $\approx$ 50 participants
- the International Conference on Ad Hoc Networks and Wireless (AdHoc-Now) from July 4th to July 6th (https://project.inria.fr/AdHocNow2016/), Attendees : academics, international $\approx$ 50 participants
- the Rescom days in January 12th and 13th, Attendees : academics, french-speaking $\approx$ 70 participants

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Valeria Loscri is co-program chairs for AdHocNow 2016.
- Nathalie Mitton is/was (co-)program chair of the RESCOM community days in January 2016, for AdHocNow 2016, UIC 2016, VTC-Spring 2016, IoTIP 2016, ICNC 2016 and CSE 2016. She is co-program chair for WSCP 2017, VTC2017 and SWC 2017.

10.1.2.2. Member of the Conference Program Committees

- Valeria Loscri is/was in the Technical Program Committee (TPC) of MoWNet'16, IDCS 2016, HPCC 2016, WiMob 2016, ICARSC 2016, GIoTS 2017, MoWNet 2017, WiMob 2017, NanoComm 2017.
- Nathalie Mitton is/was in the Technical Program Committee (TPC) ICC 2016, SmartComp 2016, GlobeCom 2016, WPMC2016, PASC 2016, Wisarn 2016, CCNC 2017, VTC 2017
- Simon Duquennoy is TPC member of EWSN 2016 (and 2017), IEEE DCoSS 2016 (and 2017), IEEE MASS 2016, InterIoT 2016, UIC 2016

10.1.3. Journal

- Valeria Loscri is Co-editor in the Special Issue on Smart Wireless Access Networks and Systems For Smart Cities, Elsevier Ad Hoc Networks.
- Valeria Loscri is Co-editor for the book MIoT2015, Management of Cyber Physical Objects in the Future Internet of Things on behalf of Springer.
- Valeria Loscri is Co-editor for the book Vehicular Social Networks, on behalf of CRC Press Taylor and Francis Group.
- Valeria Loscri is a member of the 'Research Group on IoT Communications and Networking Infrastructure' at ComSoc Communities.

10.1.3.1. Member of the Editorial Boards

- Nathalie Mitton is editorial board members of AHSWN since 2011.
- Nathalie Mitton is editorial board member of Adhoc Networks since 2012.
- Nathalie Mitton is editorial board member of IET-WSS since 2013.
- Nathalie Mitton is editorial board member of ComSoc MMTC e-letter since 2014.
- Nathalie Mitton is editorial board member of Wiley Transactions Emerging Telecommunications Technologies since 2016.
- Nathalie Mitton is editorial board member of Wireless Communications and Mobile Computing since 2016.
- Valeria Loscri is editorial board member of International Journal of Advanced Robotic Systems since 2016

10.1.4. Invited Talks

- Nathalie Mitton was invited speaker at CNRS Josy Days in Paris in October 2016
- Nathalie Mitton was invited speaker at Smyle Workshop at EPFL in Neuchatel, Switzerland, in September 2016
- Riccardo Petrolo was invited speaker at Rice University in Houston, USA in September 2016
- Nathalie Mitton was invited speaker at 4th International Workshop on Next Generation Green Wireless Networks (NextGWiN 2016) in Dublin, Ireland, September 2016
- Nathalie Mitton was invited speaker at Southeast University Seminar in Nanjing, China in June 2016
- Valeria Loscri was invited speaker at BIS days in Paris, June 2016

- Nathalie Mitton was invited speaker at 10th European eAccessibility Forum "e-Accessibility in a Connected World in Paris, May 2016
- Nathalie Mitton was invited speaker at UCN@Sophia Labex seminar Sophia Antipolis, April 2016
- Nathalie Mitton was invited speaker at "US-EU Workshop on the Next Generation Internet", Los Angeles, March 206
- Nathalie Mitton was invited speaker at "IoT day Paris Descartes" on March 2016
- Valeria Loscri and Nathalie Mitton were invited speakers at JEAI days in Lille on February 2016
- Simon Duquenoy was invited speaker at Bristol for the Sphere IRC project, in November 2016

10.1.5. Leadership within the Scientific Community

Nathalie Mitton is a member of the Steering Committee of the GDR Rescom.

10.1.6. Scientific Expertise

Nathalie Mitton is an elected member of the evaluation community of Inria. She has acted as a reviewer for ANRT and ANR project submissions. She is also member of the scientific committees of the competitiveness cluster of MATIKEN and for CITC (International Contactless Technologies Center).

Valeria Loscri is Scientific European Responsible for Inria Lille - Nord Europe. She has been reviewer in the context of 2017 Air Force Young Investigator Research Program. *Simon*

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

E-learning

Mooc, Nathalie Mitton, "Villes intelligentes : défis technologiques et sociétaux" , 5-week mooc by the IPL CityLab@Inria team, FUN, Inria, to open in January 2016

Master : Valeria Loscri, Objets Communicants, 24h (Mineure Habitat Intelligent), Ecole des Mines de Douai, France.

Master : Nathalie Mitton, Wireless sensor networks, 16h eqTD (Master MINT), Université Lille 1 and Telecom Lille 1, France

Bsc: Nathalie Mitton, Contactless technologies, 20h eqTD, Université de Valenciennes, France

Bsc: Farouk Mezghani, Contactless technologies, 20h eqTD, Université de Valenciennes, France

BSC: Riccardo Petrolo, Réseaux Informatique, 36h TP and TD (Licence Info), Université Lille 1, France

BSC: Riccardo Petrolo, Technologies du Web, 36h TP and TD (Licence 1), Université Lille 1, France

Master : Viktor Toldov, Systèmes Numériques (VHDL), 6h eqTD (Formation d'ingénieur), Télécom Lille, France

Licence : Jean Razafimandimby, Algorithms and Programming 1, 32h TP and TD, Université Lille 1, France

Licence : Jean Razafimandimby, Algorithms and Programming 2, 32h TP and TD, Université Lille 1, France

10.2.2. Supervision

PhD defended on July 26th 2016: Mouna Rekik, geographic multi path routing protocol based on swarm intelligence for wireless sensor and actuator networks in the context of Smart Grids, co-supervision Université Lille 1 and University of Sfax (Tunisia), 2013-2017, Nathalie Mitton and Zied Chtourou

PhD defended on Oct 25th 2016: Riccardo Petrolo, Internet of Things and Smart Cities, Université Lille 1, 2013-2016, Nathalie Mitton and Valeria Loscri

PhD in progress: Viktor Toldov, : Interférence et consommation dans les réseaux de capteurs, Université Lille 1, 2013-2016, Nathalie Mitton and Laurent Clavier

PhD in progress: Jean Razafimandimby, Distributed Cooperation and Communication among Heterogeneous Devices, Université Lille 1, 2014-2017, Nathalie Mitton and Valeria Loscri

PhD in progress: Aziz Mbacke, Smart Deployment of heterogeneous sensors and RFID in a Smart City, Université Lille 1, 2015-2018, Nathalie Mitton and Hervé Rivano (Urbanet)

PhD in progress: Jad Nassar, Ubiquitous networks for smart grids, Université Lille 1, 2015-2018, Nathalie Mitton and Nicolas Gouvy (HEI)

10.2.3. Juries

- PhD/HDR committees :
  - Simon Duquennoy was a member of the PhD defense committee of
    - Mathieu Michel, University Mons Belgium, September 2016
    - Anwar Hithnawi, ETH Zurich in November 2016.
  - Valeria Loscri was a member of the PhD defense committee of Riad Mazloum, University Paris 6, December 2016.
  - Valeria Loscri was reviewer of the the following PhD thesis: Orazio Briante, Universitá Mediterranea di reggio Calabria, May 2016.
  - Nathalie Mitton is/was reviewer of the following PhD thesis:
    - Remy Leone, UPMC, June 2016,
    - Quentin Bramas, UPMC, October 2016,
    - Laurent Reynaud, Univ. Lyon 1, March 2017.
  - Nathalie Mitton was a member of the HDR defense committee of Nadjib AIT SAADI, University of Paris-Est Creteil Val de Marne (UPEC) 18 july 2016.
- Researcher selection committees :
  - Valeria Loscri was member of the Inria CR2 Lille competition selection committee.
  - Nathalie Mitton was member of the Inria CR2 Grenoble competition selection committee.
  - Nathalie Mitton was member of the Inria Research Grant Position selection committee.
  - Nathalie Mitton was member of the Assistant Professor (MdC) at Universités de Valenciennes, Rennes et de Lorraine, Institut Mines Telecom ParisTech competition selection committees.
- Nathalie Mitton is a member of the scientific committee of "Convergences du Droit et du Numérique".
- Valeria Loscri is a member of the committe of "Journée Nationale de l'Internet des Objets".

10.3. Popularization

- Tahiry Razafindralambo and Simon Duquennoy gave a talk for 13-45 in April 2016
- Emilio Compagnone gave a talk for 13-45 in May 2016
- Cristina Cano gave a talk at "30 min de Sciences" in May 2016
- Emilio Compagnone and Valeria Loscri gave a talk and a demo at Robotic Day at Euratechnologies in July 2016
- Nathalie Mitton, Cristina Cano and Viktor Toldov gave talks at IoTWeek by CITC on June 2016
- Nathalie Mitton gave a talk at NormaFret by i-Trans competitiveness cluster in September 2016

- Aziz Mbacke, Jad Nassar and Julien Vandaele gave some talks in high schools for the "Fête de la Science" in October 2016
- Nathalie Mitton co-authored popularization articles in La main à la pâte (to be published in January 2017), white paper on e-accessibility for BrailleNet/G3ict and an article in *Big Data à découvert* by CNRS edition.

11. Bibliography

Major publications by the team in recent years

[1] V. Loscrí, L. Matekovits, I. Peter, A. M. Vegni.*In-body Network Biomedical Applications: from Modeling to Experimentation*, in "IEEE Transactions on NanoBioscience", 2016 [*DOI :* 10.1109/TNB.2016.2521386], https://hal.inria.fr/hal-01262020.

[2] F. Mangano, S. Duquennoy, N. Kosmatov.*Formal Verification of a Memory Allocation Module of Contiki with Frama-C: a Case Study*, in "CRiSIS 2016 - 11th International Conference on Risks and Security of Internet and Systems", Roscoff, France, September 2016, https://hal.inria.fr/hal-01351142.

[3] A. A. Mbacké, N. Mitton, H. Rivano.*Distributed Efficient & Fair Anticollision for RFID protocol*, in "IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)", New York, United States, Proc. of WiMob, October 2016, https://hal.inria.fr/hal-01351350.

[4] L. Militano, M. Erdelj, A. Molinaro, N. Mitton, A. Iera.*Recharging vs. Replacing Sensor Nodes Using Mobile Robots for Network Maintenance*, in "Telecommunications Systems", February 2016, https://hal.inria.fr/hal-01242236.

[5] K. Miranda, A. Molinaro, T. Razafindralambo.*A Survey on Rapidly Deployable Solutions for Post-disaster Networks*, in "IEEE Communications Magazine", April 2016 [*DOI :* 10.1109/MCOM.2016.7452275], https://hal.inria.fr/hal-01285617.

[6] R. Petrolo, V. Loscri, N. Mitton.*Confident-based Adaptable Connected objects discovery to HArmonize smart City Applications*, in "Proceedings of WD - IFIP Wireless Days", Toulouse, France, March 2016, https://hal.inria.fr/hal-01269633.

[7] C. Razafimandimby, V. Loscri, A. M. Vegni.*A neural network and IoT-based scheme for performance assessment in Internet of Robotic Things*, in "I4T - 1st International Workshop on Interoperability, Integration, and Interconnection of Internet of Things Systems", Berlin, Germany, April 2016, https://hal.inria.fr/hal-01261842.

[8] V. Toldov, L. Clavier, V. Loscrí, N. Mitton.*A Thompson Sampling Approach to Channel Exploration-Exploitation Problem in Multihop Cognitive Radio Networks*, in "27th annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)", Valencia, Spain, September 2016, https://hal.inria.fr/hal-01355002.

[9] V. Toldov, R. Igual-Pérez, R. Vyas, A. Boé, L. Clavier, N. Mitton.*Experimental Evaluation of Interference Impact on the Energy Consumption in Wireless Sensor Networks*, in "17th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)", Coimbra, Portugal, June 2016, https://hal.inria.fr/hal-01289487.

[10] A. M. VEGNI, V. LOSCRÍ, A. NERI, M. LEO.*A Bayesian Packet Sharing Approach for Noisy IoT Scenarios*, in "1st International Workshop on Interoperability, Integration, and Interconnection of Internet of Things Systems (I4T 2016)", Berlin , Germany, April 2016, https://hal.inria.fr/hal-01262024.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] R. PETROLO.*Semantic-based discovery and integration of heterogeneous things in a Smart City environment*, University of Lille 1, October 2016, https://hal.inria.fr/tel-01403844.

[12] M. REKIK.*Geographical multipath routing based on swarm intelligence for wireless sensorsand actuators networks: Application to Smart Grids*, Université Lille 1 ; Université de Sfax, July 2016, https://hal.inria.fr/tel-01370723.

Articles in International Peer-Reviewed Journal

[13] M. ERDELJ, N. MITTON, T. RAZAFINDRALAMBO.*Robust Wireless Sensor Network Deployment*, in "Discrete Mathematics and Theoretical Computer Science", 2016, vol. 17, n° 3, p. 105-130, https://hal.inria.fr/hal-01296574.

[14] F. GUERRIERO, V. LOSCRÍ, P. PACE, R. SURACE.*Neural Networks and SDR Modulation schemes for wireless mobile nodes: a synergic approach*, in "Ad Hoc Networks", 2017 [*DOI :* 10.1016/J.ADHOC.2016.09.016], https://hal.inria.fr/hal-01386749.

[15] V. LOSCRI, A. M. VEGNI, G. FORTINO.*On the Interaction between a Nanoparticulate System and the Human Body in Body Area Nanonetworks*, in "Micromachines", 2016, vol. 6, n° 9, p. 1213-1235 [*DOI :* 10.3390/MI6091213], https://hal.inria.fr/hal-01188280.

[16] V. LOSCRÍ, L. MATEKOVITS, I. PETER, A. M. VEGNI.*In-body Network Biomedical Applications: from Modeling to Experimentation*, in "IEEE Transactions on NanoBioscience", 2016 [*DOI :* 10.1109/TNB.2016.2521386], https://hal.inria.fr/hal-01262020.

[17] V. LOSCRÍ, A. M. VEGNI.*Capacity evaluation of a quantum-based channel in a biological context*, in "IEEE Transactions on NanoBioscience", 2017, https://hal.inria.fr/hal-01386842.

[18] A. A. MBACKÉ, N. MITTON, H. RIVANO.*RFID Reader Anticollision Protocols for Dense and Mobile Deployments*, in "MDPI Electronics Special Issue "RFID Systems and Applications"", December 2016, https://hal.inria.fr/hal-01402639.

[19] L. MILITANO, M. ERDELJ, A. MOLINARO, N. MITTON, A. IERA.*Recharging vs. Replacing Sensor Nodes Using Mobile Robots for Network Maintenance*, in "Telecommunications Systems", February 2016, https://hal.inria.fr/hal-01242236.

[20] K. MIRANDA, A. MOLINARO, T. RAZAFINDRALAMBO.*A Survey on Rapidly Deployable Solutions for Post-disaster Networks*, in "IEEE Communications Magazine", April 2016 [*DOI :* 10.1109/MCOM.2016.7452275], https://hal.inria.fr/hal-01285617.

[21] R. PETROLO, R. MORABITO, V. LOSCRÌ, N. MITTON.*The design of the gateway for the Cloud of Things*, in "Annals of Telecommunications - annales des télécommunications", May 2016 [*DOI :* 10.1007/S12243-016-0521-Z], https://hal.inria.fr/hal-01321666.

[22] T. RAZAFINDRALAMBO, M. ERDELJ, D. ZORBAS, E. NATALIZIO.*Spread and shrink: Point of interest discovery and coverage with mobile wireless sensors*, in "Journal of Parallel and Distributed Computing", November 2016 [*DOI :* 10.1016/J.JPDC.2016.09.003], https://hal.inria.fr/hal-01406492.

[23] M. REKIK, Z. CHTOUROU, N. MITTON, A. ATIEH.*Geographic routing protocol for the deployment of virtual power plant within the smart grid*, in "Sustainable Cities and Society", 2016 [*DOI :* 10.1016/J.SCS.2015.12.003], https://hal.inria.fr/hal-01252686.

[24] G. SMART, N. DELIGIANNIS, R. SURACE, V. LOSCRI, G. FORTINO, Y. ANDREOPOULOS.*Decentralized Time-Synchronized Channel Swapping for Ad Hoc Wireless Networks*, in "IEEE Transactions on Vehicular Technology", 2016, https://hal.archives-ouvertes.fr/hal-01241833.

[25] A. M. VEGNI, V. LOSCRI.*Characterization and Performance Analysis of a Chiral-Metamaterial Channel with Giant Optical Activity for Terahertz Communications*, in "Nano Communication Networks", September 2016, vol. 9, p. 28-35 [*DOI :* 10.1016/J.NANCOM.2016.07.004], https://hal.inria.fr/hal-01379904.

[26] A. M. VEGNI, V. LOSCRÍ.*Chirality effects on channel modeling for THz-band wireless communications in LoS/NLoS propagation*, in "Nano Communication Networks", July 2016, https://hal.inria.fr/hal-01345401.

[27] Y. WANG, J. ZHENG, N. MITTON.*Delivery Delay Analysis for Roadside Unit Deployment in Vehicular Ad Hoc Networks with Intermittent Connectivity*, in "IEEE Transactions on Vehicular Technology", January 2016, vol. PP, n° 99 [*DOI :* 10.1109/TVT.2015.2506599], https://hal.inria.fr/hal-01233732.

[28] T. WATTEYNE, V. HANDZISKI, X. VILAJOSANA, S. DUQUENNOY, O. HAHM, E. BACCELLI, A. WOLISZ.*Industrial Wireless IP-based Cyber Physical Systems*, in "Proceedings of the IEEE", March 2016, vol. 104, n° 5, p. 1025-1038 [*DOI :* 10.1109/JPROC.2015.2509186], https://hal.inria.fr/hal-01282597.

International Conferences with Proceedings

[29] A. DAS, A. SEN, A. MAZUMDER, N. MITTON.*On Mobile Sensor Data Collection Using Data Mules*, in "9th International Workshop on Wireless Sensor, Actuator and Robot Networks (WiSARN 2016)", Kauai, United States, International Conference on Computing, Networking and Communications – ICNC 2016 - WiSARN, February 2016, https://hal.inria.fr/hal-01212900.

[30] A. DAS, A. SEN, C. QIAO, N. GHANI, N. MITTON.*A Network Planning and Management Tool for Mitigating the Impact of Spatially Correlated Failures in Infrastructure Networks*, in "International Conference on Design of Reliable Communication Networks (DRCN)", Paris, France, March 2016, https://hal.inria.fr/hal-01254982.

[31] A. ELSTS, S. DUQUENNOY, X. FAFOUTIS, G. OIKONOMOU, R. PIECHOCKI, I. CRADDOCK.*Microsecond-Accuracy Time Synchronization Using the IEEE 802.15.4 TSCH Protocol*, in "IEEE SenseApp 2016 - Eleventh IEEE International Workshop on Practical Issues in Building Sensor Network Applications", Dubai, United Arab Emirates, November 2016, https://hal.inria.fr/hal-01351144.

[32] F. MANGANO, S. DUQUENNOY, N. KOSMATOV.*Formal Verification of a Memory Allocation Module of Contiki with Frama-C: a Case Study*, in "CRiSIS 2016 - 11th International Conference on Risks and Security of Internet and Systems", Roscoff, France, September 2016, https://hal.inria.fr/hal-01351142.

[33] A. A. MBACKÉ, N. MITTON, H. RIVANO.*Distributed Efficient & Fair Anticollision for RFID Protocol*, in "WiMob 2016 - IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications", New York, United States, October 2016, https://hal.inria.fr/hal-01351350.

[34] A. A. MBACKÉ, N. MITTON, H. RIVANO.*RFID anticollision in dense mobile environments*, in "IEEE Wireless Communications and Networking Conference", San Francisco, United States, March 2017, https://hal.inria.fr/hal-01418170.

[35] D. S. V. MEDEIROS, M. E. M. CAMPISTA, N. MITTON, M. DIAS DE AMORIM, G. PUJOLLE.*Weighted Betweenness for Multipath Networks*, in "Global Information Infrastructure and Networking Symposium (GIIS)", Porto, Portugal, October 2016, https://hal.inria.fr/hal-01370710.

[36] R. MORABITO, R. PETROLO, V. LOSCRI, N. MITTON.*Enabling a lightweight Edge Gateway-as-a-Service for the Internet of Things*, in "NOF 2016 - 7th International Conference on Network of the Future", Buzios, Rio de Janeiro, Brazil, November 2016, https://hal.inria.fr/hal-01371423.

[37] A. SEN, A. DAS, C. ZHOU, A. MAZUMDER, N. MITTON, A. A. MBACKÉ.*Reader Scheduling for Tag Access in RFID Systems*, in "NoF 2016 - 7th International Conference on Network of the Future", Buzios, Rio de Janeiro, Brazil, November 2016, https://hal.inria.fr/hal-01372697.

Conferences without Proceedings

[38] D. C. CATTARUZZA, L. BROTCORNE, N. MITTON, T. RAZAFINDRALAMBO, F. SEMET.*A two-phase matheuristic for the multi-robot routing problem under connectivity constraints*, in "Congrès annuel de la société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF)", Compiègne, France, February 2016, https://hal.inria.fr/hal-01256730.

[39] A. HITHNAWI, S. LI, H. SHAFAGH, J. GROSS, S. DUQUENNOY.*CrossZig: Combating Cross-Technology Interference in Low-Power Wireless Networks*, in "International Conference on Information Processing in Sensor Networks (ACM/IEEE IPSN 2016)", Vienna, Austria, April 2016 [*DOI :* 10.1109/IPSN.2016.7460663], https://hal.inria.fr/hal-01317464.

[40] V. LOSCRÍ, A. M. VEGNI.*Channel Modeling in a Phonon-based Quantum Network for Nano-communications*, in "NanoCom", New York, United States, ACM, September 2016 [*DOI :* 10.1145/2967446.2967480], https://hal.inria.fr/hal-01345403.

[41] A. MASKOOKI, V. TOLDOV, L. CLAVIER, V. LOSCRÍ, N. MITTON.*Competition: Channel Exploration/Exploitation Based on a Thompson Sampling Approach in a Radio Cognitive Environment*, in "EWSN - International Conference on Embedded Wireless Systems and Networks (dependability competition)", Graz , Austria, February 2016, https://hal.inria.fr/hal-01249135.

[42] R. PETROLO, S. GUZZO BONIFACIO, V. LOSCRI, N. MITTON.*The discovery of "relevant" data-sources in a Smart City environment*, in "Proceedings of SSC - 2nd International IEEE SMARTCOMP Workshop on Sensors and Smart Cities", St. Louis, Missouri, United States, May 2016, https://hal.inria.fr/hal-01300335.

[43] R. PETROLO, V. LOSCRI, N. MITTON.*Confident-based Adaptable Connected objects discovery to HArmonize smart City Applications*, in "Proceedings of WD - IFIP Wireless Days", Toulouse, France, March 2016, https://hal.inria.fr/hal-01269633.

[44] R. PETROLO, A. ROUKOUNAKI, V. LOSCRI, N. MITTON, J. SOLDATOS.*Connecting physical things to a SmartCity-OS*, in "Proceedings of CoWPER - International IEEE SECON Workshop on Toward a city-wide pervasive environment", London, United Kingdom, June 2016, https://hal.inria.fr/hal-01309638.

[45] C. RAZAFIMANDIMBY, V. LOSCRI, A. M. VEGNI, A. NERI.*A Bayesian and Smart Gateway Based Communication For Noisy IoT Scenario*, in "International Conference on Computing, Networking and Communications", Silicon Valley, United States, January 2017, https://hal.inria.fr/hal-01372546.

[46] C. RAZAFIMANDIMBY, V. LOSCRI, A. M. VEGNI.*A neural network and IoT-based scheme for performance assessment in Internet of Robotic Things*, in "I4T - 1st International Workshop on Interoperability, Integration, and Interconnection of Internet of Things Systems", Berlin, Germany, April 2016, https://hal.inria.fr/hal-01261842.

[47] *Best Paper*
V. TOLDOV, L. CLAVIER, V. LOSCRÍ, N. MITTON.*A Thompson Sampling Approach to Channel Exploration-Exploitation Problem in Multihop Cognitive Radio Networks*, in "27th annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)", Valencia, Spain, September 2016, https://hal.inria.fr/hal-01355002.

[48] V. TOLDOV, R. IGUAL-PÉREZ, R. VYAS, A. BOÉ, L. CLAVIER, N. MITTON.*Experimental Evaluation of Interference Impact on the Energy Consumption in Wireless Sensor Networks*, in "17th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)", Coimbra, Portugal, June 2016, https://hal.inria.fr/hal-01289487.

[49] V. TOLDOV, J. MEIJERS, R. IGUAL-PÉREZ, R. WOLHUTER, N. MITTON, L. CLAVIER.*Performance evaluation of LoRa radio solution for PREDNET wildlife animal tracking project*, in "LPWAN 2016", Paris Roissy, France, May 2016, https://hal.inria.fr/hal-01288077.

[50] A. M. VEGNI, V. LOSCRÍ, A. NERI, M. LEO.*A Bayesian Packet Sharing Approach for Noisy IoT Scenarios*, in "1st International Workshop on Interoperability, Integration, and Interconnection of Internet of Things Systems (I4T 2016)", Berlin , Germany, April 2016, https://hal.inria.fr/hal-01262024.

[51] A. M. VEGNI, V. LOSCRÍ.*Performance of a Chirality-affected Channel exhibiting Giant Optical Activity for Terahertz Communications*, in "NanoCom", New York, United States, ACM, September 2016 [*DOI :* 10.1145/2967446.2967453], https://hal.inria.fr/hal-01345407.

Scientific Books (or Scientific Book chapters)

[52] R. PETROLO, V. LOSCRI, N. MITTON.*Cyber-Physical Objects as key elements for a Smart Cyber-City*, in "Management of Cyber Physical Objects in the Future Internet of Things", February 2016, https://hal.inria.fr/hal-01257162.

[53] A. M. VEGNI, V. LOSCRI, A. VASILAKOS.*Vehicular Social Networks*, CRC Press , November 2016, https://hal.archives-ouvertes.fr/hal-01336988.

Scientific Popularization

[54] A. GUERRIERI, V. LOSCRI, A. ROSELLA, G. FORTINO.*Management of Cyber Physical Objects in the Future Internet of Things: Methods, Architectures and Applications*, Springer, 2016, https://hal.archives-ouvertes.fr/hal-01319193.

[55] V. LOSCRI, A. M. VEGNI, D. P. AGRAWAL.*Opportunistic Cognitive Networks*, in "Towards Cognitive Vehicular Networks", CRC Press, February 2016, https://hal.archives-ouvertes.fr/hal-01319027.

Other Publications

[56] C. CANO, D. MALONE.*A Learning Approach to Decentralised Beacon Scheduling*, May 2016, Accepted for publication in Elsevier Ad-Hoc Networks, https://hal.archives-ouvertes.fr/hal-01317207.

[57] S. DUQUENNOY, O. LANDSIEDEL, C. A. BOANO, M. ZIMMERLING, J. BEUTEL, M. C. CHAN, O. GNAWALI, M. MOHAMMAD, L. MOTTOLA, L. THIELE, X. VILAJOSANA, T. VOIGT, T. WATTEYNE.*Poster Abstract: A Benchmark for Low-power Wireless Networking*, November 2016, ACM Conference on Embedded Networked Sensor Systems (SenSys), Poster, https://hal.inria.fr/hal-01377942.

References in notes

[58] I. AMADOU, A. A. MBACKÉ, N. MITTON.*How to improve CSMA-based MAC protocol for dense RFID reader-to-reader Networks?*, in "ADHOC-NOW - 13th International Conference on Ad-Hoc Networks and Wireless", Benidorm, Spain, June 2014, https://hal.inria.fr/hal-00969117.

[59] N. ANCIAUX, L. BOUGANIM, T. DELOT, S. ILARRI, L. KLOUL, N. MITTON, P. PUCHERAL.*Opportunistic Data Services in Least Developed Countries: Benefits, Challenges and Feasibility Issues*, in "ACM SIGMOD Record", March 2014, https://hal.inria.fr/hal-00971805.

[60] J. BACHRACH, C. TAYLOR.*Localization in sensor networks*, in "Artificial Intelligence", 2004, p. 227–310.

[61] P. BOSE, P. MORIN, I. STOJMENOVIC, J. URRUTIA.*Routing with guaranteed delivery in ad hoc wireless networks*, in "Proc. of the 3rd Int. Workshop on Discrete Algorithms and Methods for Mobile Computing and Comm. (DIAL-M)", Seattle, WA, USA, August 1999, p. 48-55.

[62] H. CHAOUCHI.*Internet of Things. Connecting Objects*, Wiley and Sons, January 2010.

[63] S. DUQUENNOY, B. AL NAHAS, O. LANDSIEDEL, T. WATTEYNE.*Orchestra: Robust Mesh Networks Through Autonomously Scheduled TSCH*, in "Proceedings of the International Conference on Embedded Networked Sensor Systems (ACM SenSys)", 2015.

[64] E. ELHAFSI, N. MITTON, B. PAVKOVIC, D. SIMPLOT-RYL.*Energy-aware Georouting with Guaranteed Delivery in Wireless Sensor Networks with Obstacles*, in "International Journal of Wireless Information Networks", 2009, vol. 16, n^{o} 3, p. 142–153.

[65] B. FABIAN, O. GÜNTHER.*Distributed ONS and its Impact on Privacy*, in "IEEE International Conference on Communications", 2007, p. 1223–1228.

[66] A. GALLAIS, J. CARLE, D. SIMPLOT-RYL, I. STOJMENOVIC.*Localized Sensor Area Coverage with Low Communication Overhead*, in "IEEE Transactions on Mobile Computing (TMC)", May 2008, vol. 5, n^{o} 7, p. 661-672.

[67] N. GOUVY, N. MITTON, J. ZHENG.*Greedy Routing Recovery Using Controlled Mobility in Wireless Sensor Networks*, in "ADHOC-NOW - The 12th International Conference on Ad Hoc Networks and Wireless - 2013", Wroclaw, Poland, July 2013, http://hal.inria.fr/hal-00830374.

[68] F. INGELREST, N. MITTON, D. SIMPLOT-RYL.*A Turnover based Adaptive HELLO Protocol for Mobile Ad Hoc and Sensor Networks*, in "Proc. of the 15th IEEE International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS'07)", Bogazici University, Istanbul, Turkey, October 2007.

[69] X. LI, N. MITTON, D. SIMPLOT-RYL.*Mobility Prediction Based Neighborhood Discovery in Mobile Ad Hoc Networks*, in "Proc. of the IFIP/TC6 NETWORKING 2011", Valencia, Spain, 2011.

[70] V. LOSCRI, C. MARCHAL, N. MITTON, G. FORTINO, A. VASILAKOS V..*Security and Privacy in Molecular Communication and Networking: Opportunities and Challenges*, in "Transactions on NanoBioscience", August 2014, https://hal.inria.fr/hal-01071562.

[71] V. LOSCRI, A. M. VEGNI.*An Acoustic Communication Technique of Nanorobot Swarms for Nanomedicine Applications*, in "IEEE Transactions on NanoBioscience", April 2015, p. 1-10 [*DOI :* 10.1109/TNB.2015.2423373], https://hal.inria.fr/hal-01144003.

[72] N. MITTON, D. SIMPLOT-RYL.*Wireless Sensor and Robot Networks From Topology Control to Communication Aspects*, WorldScientific, February 2014, 284, https://hal.inria.fr/hal-00927526.

[73] A. PAL.*Localization Algorithms in Wireless Sensor Networks: Current Approaches and Future Challenges*, in "Network Protocols and Algorithms", 2010, vol. 2, n^{o} 1.

[74] R. PETROLO, T. DELOT, N. MITTON, A. MOLINARO, C. CAMPOLO.*O-SPIN: an Opportunistic Data Dissemination Protocol for Folk-enabled Information System in Least Developed Countries*, in "ADHOC-NOW - 13th International Conference on Ad-Hoc Networks and Wireless", Benidorm, Spain, June 2014, https://hal.inria.fr/hal-00960023.

[75] R. PETROLO, V. LOSCRI, N. MITTON.*Towards a smart city based on cloud of things, a survey on the smart city vision and paradigms*, in "Transactions on Emerging Telecommunications Technologies", 2015, p. 1–11 [*DOI :* 10.1002/ETT.2931], https://hal.inria.fr/hal-01116370.

[76] L. SCHMIDT, N. MITTON, D. SIMPLOT-RYL, R. DAGHER, R. QUILEZ.*DHT-based distributed ALE engine in RFID Middleware*, in "Proc. 7th IEEE International Conference on RFID-Technologies and Applications (RFID-TA)", Barcelona, Spain, 2011.

Team INOCS

Integrated Optimization with Complex Structure

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Lille - Nord Europe

THEME
Optimization, machine learning and statistical methods

Table of contents

Team INOCS

Creation of the Team: 2015 May 01

Keywords:

Computer Science and Digital Science:

6. - Modeling, simulation and control
6.1. - Mathematical Modeling
6.2. - Scientific Computing, Numerical Analysis & Optimization
6.2.3. - Probabilistic methods
6.2.6. - Optimization
8.6. - Decision support

Other Research Topics and Application Domains:

4. - Energy
4.3. - Renewable energy production
4.4. - Energy delivery
4.5. - Energy consumption
6. - IT and telecom
6.3.2. - Network protocols
7. - Transport and logistics
7.1. - Traffic management
7.1.2. - Road traffic
8.1. - Smart building/home
8.1.1. - Energy for smart buildings
8.2. - Connected city
8.4. - Security and personal assistance

1. Members

Research Scientists

Luce Brotcorne [Team leader, Inria, Researcher, HDR]
Miguel Anjos [Inria-International Chair, Researcher, HDR]

Faculty Members

Diego Cattaruzza [Ecole Centrale de Lille, Associate Professor]
Bernard Fortz [Université Libre de Bruxelles, Professor, HDR]
Martine Labbé [Université Libre de Bruxelles, Professor, HDR]
Frédéric Semet [Ecole Centrale de Lille, Professor, HDR]

Technical Staff

Maria Isabel Restrepo Ruiz [Inria]

PhD Students

Bayrem Tounsi [Inria, until Dec 2016]
Sezin Afsar [Inria, until Nov 2016]
Burak Celik [Inria/Université Libre de Bruxelles, from Nov 2016]
Yaheng Cui [Inria, from Oct 2016]
Wenjuan Gu [Ecole Centrale de Lille, from Oct 2016]

Léonard Von Niederhausern [Inria]
Yuan Yuan [Ecole Centrale de Lille, from Oct 2016]
Carlos Casorrán Amilburu [Université Libre de Bruxelles]
Jérôme de Boeck [Université Libre de Bruxelles, Teaching Assistant]
Martim Joyce-Moniz [Université Libre de Bruxelles, until Nov 2016]
Luciano Porretta [Université Libre de Bruxelles, Teaching Assistant]
Fabio Sciamannini [Université Libre de Bruxelles, Teaching Assistant]

Post-Doctoral Fellows

Hatice Calik [Université Libre de Bruxelles]
Eglantine Camby [Université Libre de Bruxelles, since Oct 2016]
Marcio Costa Santos [Université Libre de Bruxelles, since Dec 2016]
Samuel Deleplanque [Université Libre de Bruxelles]
Guillerme Duvillié [Université Libre de Bruxelles, since Oct 2016]
Diego Ponce Lopez [Université Libre de Bruxelles]

Visiting Scientists

Yasemin Arda Da Silveira [HEC-École de gestion de l'Université de Liège, from Oct 2016 until Nov 2016]
Bernard Gendron [Université de Montréal, from Oct 2016 to Dec 2016]
Juan Alejandro Gomez Herrera [Ecole Polytechnique de Montréal, Oct 2016]
Daniele Vigo [Université de Bologne, Dec 2016]

Administrative Assistants

Aurore Hermant [Inria]
Julie Vandewalle [Université Libre de Bruxelles, Secretary]

Others

Maxime Ogier [Ecole Centrale de Lille, Associate Professor]
Mohammed Skiredj [Inria, Trainee, from Mar 2016 to Aug 2016]

2. Overall Objectives

2.1. Introduction

INOCS is a cross-border "France-Belgium" project team in the Applied Mathematics Computation and Simulation Inria domain. The main goal of this team is the study of optimization problems involving complex structures. The scientific objectives of INOCS are related to modeling and methodological concerns. The INOCS team will focus on:

1. integrated models for problems with complex structure (CS) taking into account the whole structure of the problem;
2. on the development of solution methods taking explicitly into account *the nature and the structure of the decisions as well as the properties of the problem.*

Even if CS problems are in general NP-hard due to their complex nature, exact solution methods or matheuristics (heuristics based on exact optimization methods) will be developed by INOCS. The scientific contribution of INOCS will result in a toolbox of models and methods to solve challenging real life problems.

2.2. Schedule of tasks

The research program development of INOCS is to move alternatively :

- *from problems towards new approaches in optimization*: Models and solution algorithms will be developed to fit the structure and properties of the problem. From them, new generic approaches will be used to optimize problems with similar properties.
- *from innovative approaches towards problems*: The relevance of the proposed approaches will be assessed by designing new models and/or solution methods for various classes of problems. These models and methods will be based on the extension and integration of specific, well studied, models and methods.

Even if these two axes are developed sequentially in a first phase, their interactions will lead us to explore them jointly in the mid-term.

3. Research Program

3.1. Introduction

An optimization problem consists in finding a best solution from a set of feasible solutions. Such a problem can be typically modeled as a mathematical program in which decision variables must

1. satisfy a set of constraints that translate the feasibility of the solution and
2. optimize some (or several) objective function(s). Optimization problems are usually classified according to types of decision to be taken into strategic, tactical and operational problems.

We consider that an optimization problem presents a complex structure when it involves decisions of different types/nature (i.e. strategic, tactical or operational), and/or presenting some hierarchical leader-follower structure. The set of constraints may usually be partitioned into global constraints linking variables associated with the different types/nature of decision and constraints involving each type of variables separately. Optimization problems with a complex structure lead to extremely challenging problems since a global optimum with respect to the whole sets of decision variables and of constraints must be determined.

Significant progresses have been made in optimization to solve academic problems. Nowadays large-scale instances of some NP-Hard problems are routinely solved to optimality. *Our vision within INOCS is to make the same advances while addressing CS optimization problems*. To achieve this goal we aim to develop global solution approaches at the opposite of the current trend. INOCS team members have already proposed some successful methods following this research lines to model and solve CS problems (e.g. ANR project RESPET, Brotcorne *et al.* 2011, 2012, Gendron *et al.* 2009, Strack *et al.* 2009). However, these are preliminary attempts and a number of challenges regarding modeling and methodological issues have still to be met.

3.2. Modeling problems with complex structures

A classical optimization problem can be formulated as follows:

$$\begin{array}{ll} \min & f(x) \\ s.\,t. & x \in X, \end{array} \tag{1}$$

In this problem, X is the set of feasible solutions. Typically, in mathematical programming, X is defined by a set of constraints. x may be also limited to non-negative integer values.

INOCS team plan to address optimization problem where two types of decision are addressed jointly and are interrelated. More precisely, let us assume that variables x and y are associated with these decisions. A generic model for CS problems is the following:

$$\begin{array}{ll} \min & g(x,y) \\ s.\,t. & x \in X, \\ (x,y) & \in XY, \\ y & \in Y(x). \end{array} \tag{2}$$

In this model, X is the set of feasible values for x. XY is the set of feasible values for x and y jointly. This set is typically modeled through linking constraints. Last, $Y(x)$ is the set of feasible values for y for a given x. In INOCS, we do not assume that $Y(x)$ has any properties.

The INOCS team plans to model optimization CS problems according to three types of optimization paradigms: large scale complex structures optimization, bilevel optimization and robust/stochastic optimization. These paradigms instantiate specific variants of the generic model.

Large scale complex structures optimization problems can be formulated through the simplest variant of the generic model given above. In this case, it is assumed that $Y(x)$ does not depend on x. In such models, X and Y are associated with constraints on x and on y, XY are the linking constraints. x and y can take continuous or integer values. Note that all the problem data are deterministically known.

Bilevel programs allow the modeling of situations in which a decision-maker, hereafter the leader, optimizes his objective by taking explicitly into account the response of another decision maker or set of decision makers (the follower) to his/her decisions. Bilevel programs are closely related to Stackelberg (leader-follower) games as well as to the principal-agent paradigm in economics. In other words, bilevel programs can be considered as demand-offer equilibrium models where the demand is the result of another mathematical problem. Bilevel problems can be formulated through the generic CS model when $Y(x)$ corresponds to the optimal solutions of a mathematical program defined for a given x, i.e. $Y(x) = \operatorname{argmin}\{h(x,y)|y \in Y_2, (x,y) \in XY_2\}$ where Y_2 is defined by a set of constraints on y, and XY_2 is associated with the linking constraints.

In robust/stochastic optimization, it is assumed that the data related to a problem are subject to uncertainty. In stochastic optimization, probability distributions governing the data are known, and the objective function involves mathematical expectation(s). In robust optimization, uncertain data take value within specified sets, and the function to optimize is formulated in terms of a min-max objective typically (the solution must be optimal for the worst-case scenario). . A standard modeling of uncertainty on data is obtained by defining a set of possible scenarios that can be described explicitly or implicitly. In stochastic optimization, in addition, a probability of occurrence is associated with each scenario and the expected objective value is optimized.

3.3. Solving problems with complex structures

Standard solution methods developed for CS problems solve independent sub-problems associated with each type of variables without explicitly integrating their interactions or integrating them iteratively in a heuristic way. However these subproblems are intrinsically linked and should be addressed jointly. In *mathematicaloptimization* a classical approach is to approximate the convex hull of the integer solutions of the model by its linear relaxation. The main solution methods are i) polyhedral solution methods which strengthen this linear relaxation by adding valid inequalities, ii) decomposition solution methods (Dantzig Wolfe, Lagrangian Relaxation, Benders decomposition) which aim to obtain a better approximation and solve it by generating extreme points/rays. Main challenges are i) the analysis of the strength of the cuts and their separations for polyhedral solution methods, ii) the decomposition schemes and iii) the extreme points/rays generations for the decomposition solution methods.

The main difficulty in solving *bilevel problems* is due to their non convexity and non differentiability. Even linear bilevel programs, where all functions involved are affine, are computationally challenging despite their apparent simplicity . Up to now, much research has been devoted to bilevel problems with linear or convex follower problems. In this case, the problem can be reformulated as a single-level program involving complementarity constraints, exemplifying the dual nature, continuous and combinatorial, of bilevel programs.

4. Application Domains

4.1. Energy

In energy, the team mainly focuses on pricing models for demand side management. Demand side management methods are traditionally used to control electricity demand which became quite irregular recently and resulted in inefficiency in supply. We have explored the relationship between energy suppliers and customers who are connected to a smart grid. The smart grid technology allows customers to keep track of hourly prices and shift their demand accordingly, and allows the provider to observe the actual demand response to its pricing strategy. We tackle pricing problems in energy according to the bilevel optimization approaches. Some research works in this domain are supported by bilateral grants with EDF.

4.2. Transportation and Logistics

In transportation and logistics, the team addresses mainly integrated problems, which require taking into account simultaneously different types of decision. Examples are location and routing, inventory management and routing or staff scheduling and warehouse operations management. Such problems occur from the supply chain design level to the logistic facility level. Some research works in this application domain are supported by bilateral grants/contrats with Colisweb, INFRABEL or DHL.

4.3. Telecommunications

In telecommunications, the team mainly focuses on network design problems and on routing problems. Such problems are optimization problems with complex structure, since the optimization of capacity installation and traffic flow routing have to be addressed simultaneously. Some research works are conducted within a long-term cooperation with Nokia (formerly Alcatel-Lucent Bell Labs).

5. Highlights of the Year

5.1. Highlights of the Year

- Creation of the Inria Innovation : Colinocs between Colisweb (start-up devoted to attended delivery service within the next 2 hours) and INOCS.
- Miguel Anjos joined us in September as part of the Inria International Chair program and will spend 20% of his time with us until 2020.
- A joint team between Ecole des Mines de St Etienne and INOCS involving N. Absi, D. Cattaruzza, D. Feillet, M. Ogier, F. Semet was finalist of the EURO/ROADEF Challenge 2016 devoted to an Inventory Routing Problem proposed by Air Liquid.

6. New Results

6.1. Large scale complex structure optimization

New decomposition methods for the time-dependent combined network design and routing problem: A significant amount of work has been focussed on the design of telecommunication networks. The performance of different Integer Programming models for various situations has been computationally assessed. One of the settings that has been thoroughly analyzed is a variant where routing decisions (for time-dependent traffic demand), and network design, are combined in a single optimization model. Solving this model with a state-of-the-art solver on representative network topologies, shows that this model quickly becomes intractable. With an extended formulation, both the number of continuous flow variables and the number of fixed charge capacity constraints are multiplied by a factor $|V|$ (where V represents the set of nodes) leading to large model. However, the linear relaxation of this extended formulation yields much better lower bounds. Nevertheless, even if the extended model provides stronger lower bounds than the aggregated formulation, it suffers from its huge size: solving the linear relaxation of the problem quickly becomes intractable when the network size increases, making the linear relaxation expensive to solve. This observation motivates the analysis of decomposition methods [30].

Convex piecewise linear unsplittable multicommodity flow problems We studied the multi-commodity flow problem with unsplittable flows, and piecewise-linear costs on the arcs. They show that this problem is NP-hard when there is more than one commodity. We propose a new MILP models for this problem, that was compared to two formulations commonly used in the literature. The computational experiments reveal that the new model is able to obtain very strong lower bounds, and is very efficient to solve the considered problem [40].

Tree Reconstruction Problems: We studied the problem of reconstructing a tree network by knowing only its set of terminal nodes and their pairwise distances, so that the reconstructed network has its total edge weight minimized. This problem has applications in several areas, namely the inference of phylogenetic trees and the inference of routing networks topology. Phylogenetic trees allow the understanding of the evolutionary history of species and can assist in the development of vaccines and the study of biodiversity. The knowledge of the routing network topology is the basis for network tomography algorithms and it is a key strategy to the development of more sophisticated and ambitious traffic control protocols and dynamic routing algorithms [31].

Comparison of formulations and solution methods for the discrete ordered p-median problem: We presented several new formulations for the Discrete Ordered Median Problem (DOMP) based on its similarity with some scheduling problems. Some of the new formulations present a considerably smaller number of constraints to define the problem with respect to some previously known formulations. Furthermore, the lower bounds provided by their linear relaxations improve the ones obtained with previous formulations in the literature even when strengthening is not applied. We also present a polyhedral study of the assignment polytope of our tightest formulation showing its proximity to the convex hull of the integer solutions of the problem. Several resolution approaches, among which we mention a branch and cut algorithm, are compared. Extensive computational results on two families of instances, namely randomly generated and from Beasley's OR-library, show the power of our methods for solving DOMP [34].

New models and algorithms for integrated vehicle routing problems
We address a real-life inventory routing problem, which consists in designing routes and managing the inventories of the customers simultaneously. The problem was introduced during the 2016 ROADEF/EURO challenge. The proposed problem is original and complex for several reasons : the logistic ratio optimization objective, the hourly time-granularity for inventory constraints, the driver/trailer allocation management. Clearly, this problem is an optimization problem with complexe structure, for which we proposed a branch-cut-and-price based method : a cut and-column generation procedure was developed, along with a heuristic pricing algorithm to generate new columns and a heuristic fixing procedure to generate integer solutions. The solution method allowed the team including INOCS members to qualify to the final phase of the ROADEF/EURO challenge 2016 [41].

Column generation approach for pure parsimony haplotyping: The knowledge of nucleotides chains that compose the double DNA chain of an individual has a relevant role in detecting diseases and studying populations. However, determining experimentally the single nucleotides chains that, paired, form a certain portion of the DNA is expensive and time-consuming. Mathematical programming approaches have been proposed instead, e.g. formulating the Haplotype Inference by Pure Parsimony problem (HIPP). Abstractly, we are given a set of genotypes (strings over a ternary alphabet 0, 1, 2) and we want to determine the smallest set of haplotypes (binary strings over the set 0, 1) so that each genotype can be 'generated' by some pair of haplotypes, meaning that they are compatible with the genotype and can fully explain its structure. In order to deal with larger instances, we proposed a new model involving an exponential number of variables to be solved via column generation, where variables are dynamically introduced into the model by iteratively solving a pricing problem. We compared different ways of solving the pricing problem, based on integer programming, smart enumeration and local search heuristic. The efficiency of the approach is improved by stabilization and by a heuristic to provide a good initial solution. Results show that, with respect to the linear relaxations of both the polynomial and exponential-size models, our approach yields a tighter formulation and outperforms in both efficiency and effectiveness the previous model for instances with a large number of genotypes [39].

6.2. Bilevel Programming

Bilevel approaches for energy management problems: We have proposed the first bilevel pricing models to explore the relationship between energy suppliers and customers who are connected to a smart grid. Due to their definition, bilevel models enable to integrate customer response into the optimization process of supplier who aims to maximize revenue or minimize capacity requirements. In our setting, the energy provider acts as a leader (upper level) that takes into account a smart grid (lower level) that minimizes the sum of users' disutilities. The latter bases its decisions on the hourly prices set by the leader, as well as the schedule preferences set by the users for each task. The pricing problems, we model, belong to the category of single leader single follower problems. Considering both the monopolistic and competitive environment we present two bilevel bilinear bilinear problems with continuous variables. Heuristics solutions methods are defined to solve large size instances of the models. They are based on the interactions between prices, schedules and peaks. Numerical results on randomly generated instances illustrate numerically the validity of the approach, which achieves an ‘optimal trade-off between three objectives: revenue, user cost, and peak demand. Moreover, they put into highlight the ability of the heuristics to produce high quality results compared to the solution of MIP reformulations of the models[36].

New formulations for solving Stackelberg games: We analyzed general Stackelberg games (SGs) and Stackelberg security games (SSGs). SGs are hierarchical adversarial games where players select actions or strategies to optimize their payoffs in a sequential manner. SSGs are a type of SGs that arise in security applications, where the strategies of the player that acts first consist in protecting subsets of targets and the strategies of the followers consist in attacking one of the targets. We review existing mixed integer optimization formulations in both the general and the security setting and present new formulations for the the second one. We compare the SG formulations and the SSG formulations both from a theoretical and a computational point of view. We indentify which formulations provide tighter linear relaxations and show that the strongest formulation for the security version is ideal in the case of one single attacker. Our computational experiments show that the new formulations can be solved in shorter times [46].

6.3. Robust/Stochastic programming

Decomposition method for stochastic staff management problems : We addressed an integrated shift scheduling and load assignment optimization problem for attended home delivery, which is a last-mile delivery service requiring the presence of the customer for the delivery. We were mainly interested in generating a daily master plan for each courier. We proposed a tactical problem integrating a shift scheduling problem and a load assignment problem under demand uncertainty, which was modeled as a two-stage stochastic programming model. This model integrates two types of decisions. First-stage decisions are related to the design of a schedule that includes the periods of the day in which each courier must work and the o-d pairs to visit at each time period. Second-stage decisions (recourse actions) consist of the allocation of a number of packages to be delivered at each time period, for each o-d pair, by each courier, such that the demand (number of packages to deliver) for each scenario is satisfied. Recourse is the ability to take corrective actions after a random event has taken place. The objective is to minimize the sum of the daily staffing cost plus the expected daily recourse cost. To solve this problem, we proposed and implemented a multi-cut integer L-shaped algorithm, where the second stage decomposes by time periods and by demand scenarios. To strengthen the first stage model, some valid inequalities are added, and some of the existing constraints are lifted. Results on real-world based instances from a delivery company demonstrate that our approach provides robust tactical solutions that easily accommodate to fluctuations in customer orders, preventing additional costs related to the underutilization of couriers and the use of external couriers to satisfy all delivery requests [37], [43].

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

- Fluxys (2016-2018). Study of optimization problems arising in the management of gas networks.
- Colisweb (2015-2016). Study of optimization problems arising in courier scheduling. This bilateral contract leads to the creation of an Inria Innovation Lab at the end of 2016.

7.2. Bilateral Grants with Industry

- PARROT (Planning Adapter performing ReRouting and Optimization of Timing), part of BEWARE Fellowships Academia funded by the COFUND program of the European Union (FP7 - Marie Curie Actions). INFRABEL is the industrial partner of this project.(2014-2018)
- Design and Pricing of Electricity Services in a Competitive Environment within the Gaspard Monge Research Progam (PGMO) funded by the Fondation Mathématiques Jacques Hadamard. EDF is the industrial partner (2015-2018).
- BENMIP: A generic bender decomposition-based (mixed) integer programming solver within the Gaspard Monge Research Progam (PGMO) funded by the Fondation Mathématiques Jacques Hadamard.(2015-2017)

8. Partnerships and Cooperations

8.1. Regional Initiatives

The ELSAT research program addresses the issues involved in sustainable transportation and mobility. Within ELSAT, INOCS is involved on two projects devoted to hybrid optimization methods in logistics and to city logistics in collaboration with LAMIH (University of Valenciennes), LGI2A (University of Artois) and LEOST (IFSTTAR). ELSAT is supported by the CPER 2015-2020 (State-Region Contract).

8.2. National Initiatives

8.2.1. ANR

ANR project PI-Commodality “Co-modal freight transportation chains: an approach based on physical internet” in collaboration with CGS-ARMINES (Paris), LAAS (Toulouse), DHL (2016 - 2018). The PI-co-modality project aims to design new sustainable logistic services between preset origins and destinations. It is based on innovative approaches both in terms of: 1) Logistics and transportation services : by considering the PI-internet approach, specifically: mesh logistics and transportation networks based on available capacities, by designing consistent integrated co-modal chains; 2)Methodology : by addressing the underlying problems according to two approaches: centralized and decentralized, by proposing news realistic models relevant for practitioner taking into account the consistency, by developing state-of-the-art decision making algorithms.

8.2.2. National Initiatives (Belgium)

Combinatorial Optimization: Meta-heuristics and Exact Methods (2012-2017, coordinator: Bernard Fortz (GOM-ULB/INOCS-Inria). Inter-university Attraction Pole funded by the Belgian Federal Science Policy Office. Study and modeling of combinatorial optimization problems; Advancements in algorithmic techniques; Implementation of solution methods for large-scale, practically relevant problems.

8.3. European Initiatives

8.3.1. Collaborations in European Programs, Except FP7 & H2020

Program: COST

Project acronym: TD1207

Project title: Mathematical Optimization in the Decision Support Systems for Efficient and Robust Energy Networks

Duration: 04/2014 - 04/2017

Coordinator: Thorsten Koch (ZIB, Germany)

INOCS partners: Bernard Fortz, Martine Labbé

Abstract: Energy Production and Distribution (EP&D) is among the biggest challenges of our time, since energy is a scarce resource whose efficient production and fair distribution is associated with many technical, economical, political and ethical issues like environmental protection and people health. EP&D networks have rapidly increased their size and complexity, e.g. with the introduction and interconnection of markets within the EU. Thus, there is an increasing need of systems supporting the operational, regulatory and design decisions through a highly inter-disciplinary approach, where experts of all the concerned fields contribute to the definition of appropriate mathematical models. This is particularly challenging because these models require the simultaneous use of many different mathematical optimization tools and the verification by experts of the underlying engineering and financial issues. The COST framework is instrumental for this Action to be able to coordinate the inter-disciplinary efforts of scientists and industrial players at the European level.

Program: JPI Urban Europe

Project acronym: e4-share

Project title: Models for Ecological, Economical, Efficient, Electric Car-Sharing

Duration: 10/2014 - 09/2017

Coordinator: Markus Leitner (University of Vienna, Austria)

Other partners:

- Austrian Institute of Technology, Austria
- Université Libre de Bruxelles (INOCS), Belgium
- University of Bologna, Italy
- tbw research GesmbH, Austria

Abstract: Car-sharing systems and the usage of electric cars become increasingly popular among urban citizens. Thus, providing vast opportunities to meet today's challenges in terms of environmental objectives, sustainability and living quality. Our society needs to manage a transformation process that ultimately shall lead to fewer emissions and less energy consumption while increasing the quality of public space available. In e4-share, the team will lay the foundations for efficient and economically viable electric car-sharing systems by studying and solving the optimization problems arising in their design and operations. A main goal is to derive generic methods and strategies for optimized planning and operating in particular for flexible variants which best meet preferences of customers but impose nontrivial challenges to operators. This project will develop novel, exact and heuristic, numerical methods for finding suitable solutions to the optimization problems arising at the various planning levels as well as new, innovative approaches considering these levels simultaneously.

8.4. International Initiatives

8.4.1. Inria International Partners

8.4.1.1. Informal International Partners

- Department of Statistics and Operations Research, University of Vienna, Austria.
- Centre for Quantitative Methods and Operations Management, HEC-Liège, Belgique.
- Interuniversity Centre on Entreprise Networks, Transportation and Logistics, Montreal, Canada.
- Instituto Sistemas Complejos de Ingeniería (ISCI), Santiago, Chile.
- The Centre for Business Analytics, University College Dublin, Ireland.
- Department of Electrical, Electronic, and Information Engineering, University of Bologna, Italy.
- Department of Mathematics, University of Aveiro, Portugal.
- Department of Statistics and Operations Research, University of Lisbon, Portugal.

- Instituto de Matemáticas, University of Seville.
- Dipartimento di Matematica, Universita degli studi di Padova.

8.4.2. Participation in Other International Programs

- STIC Algérie, University of Oran, Algeria.

8.5. International Research Visitors

8.5.1. Visits of International Scientists

- Yasemin Arda Da Silveira, HEC-École de gestion de l'Université de Liège, Visiting Scientist from Oct 2016 until Nov 2016
- Bernard Gendron, Université de Montréal, Visiting Scientist from Oct 2016 to Dec 2016
- Juan Alejandro Gomez Herrera, Ecole Polytechnique de Montréal, Visiting Scientist Oct 2016
- Daniele Vigo, Université de Bologne, Visiting Scientist, Dec 2016.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

Luce Brotcorne:

– Meeting of the EURO Working group on Pricing and Revenue Management, Hamburg, Germany, April 2016.

Bernard Fortz:

– Winter School on Network Optimization, Estoril, Portugal, January 2016.

9.1.2. Scientific Events Selection

9.1.2.1. Chair of Conference Program Committees

Bernard Fortz:

– INFORMS Telecommunications conference, Boca Raton, FL, USA, March 2016.

9.1.2.2. Member of the Conference Program Committees

Luce Brocorne:

– ROADEF 2016, Compiègne, France, February 2016.
– EURO 2016, Stream Organizer, Poznan, Poland, July 2016.

Bernard Fortz:

– ORBEL 30, Louvain-la-Neuve, Belgium, January 2016.

Martine Labbé:

– DRCN, Paris, France, March 2016.
– International Symposium on Combinatorial Optimization (ISCO), Vietri sul Mare, Italy, May 2016.
– Ninth Triennal Symposium on transportation analysis (TRISTAN IX), June 2016.
– Meeting of the EURO working group on Locational Decisions (EWGLA), Malaga, Spain, September 2016.
– XLVIII Brazilian Symposium on Operational Research (XLVIII SBPO), Vitória, Brazil, September 2016.
– Matheuristics 2016, Brussels, Belgium, September 2016.

Frédéric Semet:

– Ninth Triennal Symposium on transportation analysis (TRISTAN IX), June 2016.
– National Conference of the Tunisian Operations Research Society (TORS), December 2016.

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

Luce Brotcorne:

- Associate editor: Computers and Operations Research

Bernard Fortz:

- Associate editor: INFORMS Journal on Computing
- Guest editor of special issues of Networks and EURO Journal on Computational Optimization

Martine Labbé:

- Editor in chief: EURO Journal on Computational Optimization
- Associate editor: International Transactions in Operations Research
- Member of the Advisory Board: Transportation Science

9.1.4. Invited Talks

Luce Brotcorne:

- CESO 2016, Plenary speaker, Paris, France, May 2016
- EDF Lab Seminar, Paris, France, France, September 2016
- RIM seminar, Erasmus, Rotterdam, Netherlands, December 2016.

Diego Cattaruzza:

- Invited seminar, HEC-École de gestion de l'université de Liège, Liège, Belgium, February 2016.

Bernard Fortz:

- OMOR seminar, ESSEC, Cergy, France, December 2016.
- SDN day 2016, Orange Gardens, Paris, France, November 2016.

Martine Labbé:

- Winter School on Network Optimization, Invited Lecturer, Lisbon, Portugal, January 2016.
- First International Workshop in Bilevel Programming, Monterrey, Plenary Speaker, Mexico, March 2016.
- ROADEF Conférence, Plenary Speaker, Compiègne, France, February 2016.
- European Study Group with Industry, Plenary Speaker, Avignon, France, May 2016
- Graphs and Optimization (GO) Meeting, Plenary Speaker, Rigi Kaldbad, Switzerland, July 2016.
- Séminaire POC15, Plenary Speaker, Paris, France, October 2016 .

Frédéric Semet

- Symposium in honor of G. Laporte, Eindhoven, Netherlands, April 2016.
- AIRO Conference, Plenary Speaker, Trieste, Italy, September 2016.

9.1.5. Leadership within the Scientific Community

Luce Brotcorne:

- Coordinator of EURO Working Group: "Pricing and Revenue Management".

Bernard Fortz:

- Member of the board of administration and treasurer of ORBEL (Belgian OR Society).
- ORBEL representative for EURO and IFORS.

– Coordinator of EURO Working Group: “European Network Optimization Group (ENOG)”.

Martine Labbé:

– Vice-chair of the SIAM Activity Group on Optimization (SIAG/OPT).
– Chair of the SIAG/Optimization Prize committee.

Frédéric Semet:

– Member of the board of EURO Working Group:“Vehicle routing and logistics optimization (VEROLOG)”.
– Member of the steering committee of CNRS GdR 3002 : Operations Research.
– Coordinator of GdR Working Group: “Transportation and Logistics (GT2L)”.

9.1.6. Scientific Expertise

Luce Brotcorne:

– Member of the scientific committee of France-Netherlands Exchange Program.
– Member of the evaluation committee for Inria/MITACS Exchange Program.

Bernard Fortz:

– President of the FRIA PE1 - jury 1.
– Member of the CIRRELT scientific orientation committee.

Martine Labbé:

– Member of the Scientific Advisory Board of IWR and its Graduate school HGS MathComp, Heidelberg University.
– Member of the Centro de Matemática, Aplicaçoes Fundamentais e Investigaçao Operacional, University of Lisbon.
– Member of the 2016 selection jury for the research program “Mathematics and ...” of the Vienna Science and Technology Fund.

Frederic Semet:

– Member of the CIRRELT scientific orientation committee.
– Scientific board member of PICOM competitiveness cluster.
– Reviewer for Agence Nationale de la Recherche (ANR), Fond de Recherche Nature et Technologie du Québec.

9.1.7. Research Administration

Luce Brotcorne:

– Scientific Manager (correspondant scientifique) for international relations department
– Member of the International Relations working (COST-GTRI).
– Member of the committee for the Technological Development (CDT).
– Member of the committee for the recruitment of Junior Research Scientist (CR1/CR2) at Inria Bordeaux and Inria Lille in 2016
– Member of the committee for the recruitment of assistant professor at University of Valenciennes in 2016

Frédéric Semet:

– Deputy director of CRIStAL.
– Elected member of the scientific council of Centrale Lille

9.2. Teaching - Supervision - Juries

9.2.1. Supervision

PhD : Lijuan Zhang, Optimisation and Simulation of a cross-dock facility, 18/03/2016, Frédéric Semet, Benoit Trouillet

PhD : Diego Ponce Lopez, The Discrete Ordered Median Problem revisited: new formulations, properties and algorithms, Université Libre de Bruxelles, 18/07/2016, Martine Labbé, Justo Puerto

PhD : Martim Joyce Moniz, Models and methods for Traffic Engineering problems with single-path routing, Université Libre de Bruxelles, 06/10/2016, Bernard Fortz, Luis Gouveia

PhD : Sezin Afsar, Revenue Optimization and Demand Response Models using Bilevel Progamming in Smart Grid Systems, 07/12/2016, Luce Brotcorne, Gilles Savard

PhD : Bayrem Tounsi, Contributions in E-commerce supply chain : Integration in E-fullfilment and delivery services pricing,19/12/2016, Luce Brotcorne, Yezekael Hayel

PhD : Kacem Danach, Hyperheuristics in logistics 21/12/2016, Shahin Gelareh, Frédéric Semet

PhD in progress : Burak Celik, Models and methods for Stackelberg games using bilevel optimization and mixed integer linear programming, from Nov 2016, Luce Brotcorne, Martine Labbé

PhD in progress : Yaheng Cui, Models and methods for decentralized decision in logistics networks, from Oct 2016, Luce Brotcorne, Eric Ballot

PhD in progress : Wenjuan Gu, Location routing for short and local fresh food supplu chain, from Oct 2016, Maxime Ogier, Frédéric Semet

PhD in progress : Léonard Von Niederhausern, Design and pricing of new services in energy in a competitive environment, from Oct 2015, Luce Brotcorne, Didier Aussel

PhD in progress : Yuan Yuan, Vehicle Routing Problems with Synchronization for City Logistics, from Oct 2016, Diego Cattaruzza, Frédéric Semet

PhD in progress : Carlos Casorrán Amilburu, Models and algorithms for Solving Bimatrix Stackelberg games, from October 2014, Martine Labbé.

PhD in progress : Jérôme De Boeck, Optimization problems in energy, from October 2015, Bernard Fortz.

PhD in progress : Luciano Porretta, Models and methods for the study of genetic associations, from May 2011, Bernard Fortz.

PhD in progress : Fabio Sciamannini, Comumn generatin approaches for solving variants of node coloring problems, from October 2014, Bernard Fortz, Martine Labbé.

9.2.2. Juries

Luce Brotcorne:

- PhD : "Design, Planning and Execution of Sustainable Intermodal Port-Hinterland Transport Networks", Ypsilantis Panagiotis, Erasmus University, Rotterdam. Rob Zuidwijk.

Bernard Fortz:

- PhD : "Revenue Optimization and Demand Response Models using Bilevel Progamming in Smart Grid Systems", Sezin Afşar, Inria Lille-Nord Europe. Luce Brotcorne and Gilles Savard.
- HdR : "Problèmes d'optimisation en milieu urbain : modèles, méthodes et défis", Andréa Cynthia Santos, Université de Technologie de Troyes.
- PhD : "The discrete ordered median problem revisited: new formulations, properties and algorithms", Diego Ponce, Université Lire de Bruxelles and Université de Séville. Martine Labbé and Justo Puerto.

– PhD : “Optimization of information flows in telecommunication networks” (rapporteur), Thibaut Lefebvre, CNAM. Sourour Elloumi, Eric Gourdin, Cédric Bentz.

Martine Labbé:

– PhD : “Recherche de flots stables dans des réseaux de transport multi-agents” (rapporteur), Nadia Chaabane, IUniversité de Toulouse. Cyril Briant and Marie-José Huguet.

– PhD : “Models and methods for Traffic Engineering problems with single-path routing”, Martim Joyce Moniz, Université Libre de Bruxelles, Bernard Fortz and Luis Gouveia.

– HdR : “Network Optimization: Algorithmic Approaches and Polyhedral Investigations:”, Markus Leitner, University of Vienna.

Frédéric Semet:

– PhD : “Design, Planning and Execution of Sustainable Intermodal Port-Hinterland Transport Networks”, Juliette Médina, Ecole des Mines de Nantes. Fabien Le Huédé, Olivier Peton.

9.3. Popularization

- PICOM workshop on Logistics, May 2016.
- Rendez-vous du Plateau Meetings: Prescriptive analytics for an agile logistics, December 2016.

10. Bibliography

Major publications by the team in recent years

[1] I. BILEGAN, L. BROTCORNE, D. FEILLET, Y. HAYEL. *Revenue Management for rail container transportation*, in "Euro Journal on Transportation and Logistics", 2015, vol. 4, p. 261-283.

[2] S. BINART, P. DEJAX, M. GENDREAU, F. SEMET. *A 2-stage method for a field service routing problem with stochastic travel and service times*, in "Computers & Operations Research", 2016, vol. 65, p. 64–75.

[3] Q. BOTTON, B. FORTZ, L. GOUVEIA. *On the hop-constrained survivable network design problem with reliable edges*, in "Computers & Operations Research", 2015, vol. 64, n° 0, p. 159 - 167 [*DOI :* 10.1016/J.COR.2015.05.009], http://www.sciencedirect.com/science/article/pii/S0305054815001264.

[4] Q. BOTTON, B. FORTZ, L. GOUVEIA, M. POSS. *Benders Decomposition for the Hop-Constrained Survivable Network Design Problem*, in "INFORMS Journal on Computing", 2013, vol. 25, n° 1, p. 13-26 [*DOI :* 10.1287/IJOC.1110.0472], http://joc.journal.informs.org/content/25/1/13.abstract.

[5] L. BROTCORNE, F. CIRINEI, P. MARCOTTE, G. SAVARD. *A tabu search algorithm for the network pricing problem*, in "Computers & Operations Research", 2012, vol. 39, n° 11, p. 2603-2611.

[6] L. BROTCORNE, S. HANAFI, R. MANSI. *One-level reformulation of the bilevel knapsack problem using dynamic programming*, in "Discrete Optimization", 2013, vol. 10, n° 1, p. 1-10.

[7] C. CASORRÁN, B. FORTZ, M. LABBÉ, F. ORDÓÑEZ. *Novel formulations for general and security Stackelberg games*, 2016, working paper or preprint, https://hal.inria.fr/hal-01429265.

[8] L. CASTELLI, M. LABBÉ, A. VIOLIN.*A network pricing formulation for the revenue maximization of european air navigation service providers*, in "Transportation Research Part C: Emerging Technologies", 2013, vol. 33, p. 214–226.

[9] D. CATANZARO, L. GOUVEIA, M. LABBÉ.*Improved integer linear programming formulations for the job Sequencing and tool Switching Problem*, in "European Journal of Operational Research", 2015, vol. 244, n° 3, p. 766–777.

[10] D. CATANZARO, M. LABBÉ, R. PESENTI, J.-J. SALAZAR-GONZÁLEZ.*The balanced minimum evolution problem*, in "INFORMS Journal on Computing", 2012, vol. 24, n° 2, p. 276–294.

[11] D. CATTARUZZA, N. ABSI, D. FEILLET, J. GONZÁLEZ-FELIU.*Vehicle routing problems for city logistics*, in "EURO Journal on Transportation and Logistics", 2015, p. 1–29 [*DOI :* DOI:10.1007/S13676-014-0074-0].

[12] D. CATTARUZZA, N. ABSI, D. FEILLET, T. VIDAL.*A memetic algorithm for the multi trip vehicle routing problem*, in "European Journal of Operational Research", 2014, vol. 236, n° 3, p. 833–848 [*DOI :* 10.1016/J.EJOR.2013.06.012].

[13] D. CATTARUZZA, N. ABSI, D. FEILLET, D. VIGO.*An iterated local search for the multi-commodity multi-trip vehicle routing problem with time windows*, in "Computers & Operations Research", 2014, vol. 51, p. 257–267 [*DOI :* 10.1016/J.TRPRO.2015.03.027].

[14] T. R. CHRISTENSEN, M. LABBÉ.*A branch-cut-and-price algorithm for the piecewise linear transportation problem*, in "European Journal of Operational Research", 2015, vol. 245, n° 3, p. 645–655.

[15] É. DUCHENNE, G. LAPORTE, F. SEMET.*The Undirected m-Capacitated Peripatetic Salesman Problem*, in "European Journal of Operational Research", 2012.

[16] B. FORTZ, M. LABBÉ, F. LOUVEAUX, M. POSS.*Stochastic binary problems with simple penalties for capacity constraints violations*, in "Mathematical Programming", 2013, vol. 138, p. 199-221, http://dx.doi.org/10.1007/s10107-012-0520-4.

[17] S. GELAREH, R. N. MONEMI, F. SEMET, G. GONCALVES.*A branch-and-cut algorithm for the truck dock assignment problem with operational time constraints*, in "European Journal of Operational Research", 2016, vol. 249, n° 3, p. 1144–1152.

[18] B. GENDRON, P.-V. KHUONG, F. SEMET.*Multilayer variable neighborhood search for two-level uncapacitated facility location problems with single assignment*, in "Networks", 2015, vol. 66, n° 3, p. 214–234.

[19] B. GENDRON, P.-V. KHUONG, F. SEMET.*A Lagrangian-based branch-and-bound algorithm for the two-level uncapacitated facility location problem with single-assignment constraints*, in "Transportation Science", 2016.

[20] J. GONZALEZ-FELIU, F. SEMET, J. ROUTHIER.*Sustainable urban logistics: concepts, methods and information systems*, Springer, 2014.

[21] N. JOZEFOWIEZ, G. LAPORTE, F. SEMET.*A generic branch-and-cut algorithm for multiobjective optimization problems: Application to the multilabel traveling salesman problem*, in "INFORMS Journal on Computing", 2012, vol. 24, n$^{\text{o}}$ 4, p. 554–564.

[22] R. LAHYANI, L. C. COELHO, M. KHEMAKHEM, G. LAPORTE, F. SEMET.*A multi-compartment vehicle routing problem arising in the collection of olive oil in Tunisia*, in "Omega", 2015, vol. 51, p. 1–10.

[23] R. LAHYANI, M. KHEMAKHEM, F. SEMET.*Rich vehicle routing problems: From a taxonomy to a definition*, in "European Journal of Operational Research", 2015, vol. 241, n$^{\text{o}}$ 1, p. 1–14.

[24] D. PAPADIMITRIOU, B. FORTZ.*A Rolling Horizon Heuristic for the Multiperiod Network Design and Routing Problem*, in "Networks", 2015, vol. 66, n$^{\text{o}}$ 4, p. 364–379, http://dx.doi.org/10.1002/net.21659.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[25] M. JOYCE-MONIZ.*Models and methods for Traffic Engineering problems with single-path routing*, Université Libre de Bruxelles (U.L.B.), Belgium, October 2016, https://hal.inria.fr/tel-01421865.

[26] D. PONCE LOPEZ.*The Discrete Ordered Median Problem revisited: new formulations, properties and algorithms*, Université Libre de Bruxelles (U.L.B.), Belgium, July 2016, https://hal.inria.fr/tel-01421868.

Articles in International Peer-Reviewed Journal

[27] S. BINART, P. DEJAX, M. GENDREAU, F. SEMET.*A 2-stage method for a field service routing problem with stochastic travel and service times*, in "Computers and Operations Research", 2016, vol. 65, p. 64-75, https://hal.archives-ouvertes.fr/hal-01223020.

[28] L. CASTELLI, M. LABBÉ, A. VIOLIN.*Network pricing problem with unit toll*, in "Networks", 2016, vol. 69, n$^{\text{o}}$ 1, p. 83–93, https://hal.inria.fr/hal-01419552.

[29] D. C. CATTARUZZA, N. ABSI, D. FEILLET.*Vehicle routing problems with multiple trips*, in "4OR: A Quarterly Journal of Operations Research", 2016, https://hal-emse.ccsd.cnrs.fr/emse-01250603.

[30] B. FORTZ, E. GORGONE, D. PAPADIMITRIOU.*A Lagrangian heuristic algorithm for the time-dependent combined network design and routing problem*, in "Networks", 2017, vol. 69, n$^{\text{o}}$ 1, p. 110–123 [*DOI :* 10.1002/NET.21721], https://hal.archives-ouvertes.fr/hal-01428061.

[31] B. FORTZ, O. OLIVEIRA, C. REQUEJO.*Compact Mixed Integer Linear Programming models to the Minimum Weighted Tree Reconstruction Problem*, in "European Journal of Operational Research", 2017, vol. 256, n$^{\text{o}}$ 1, p. 242 - 251, https://hal.inria.fr/hal-01410554.

[32] S. GELAREH, R. N. MONEMI, F. SEMET, G. GONCALVES.*A branch-and-cut algorithm for the truck dock assignment problem with operational time constraints*, in "European Journal of Operational Research", 2016, https://hal.inria.fr/hal-01429317.

[33] M. LABBÉ, A. MARIN, A. M. RODRIGUEZ-CHIA.*Lexicographical Order in Integer Programming*, in "Vietnam Journal of Mathematics", 2016 [*DOI :* 10.1007/s10013-016-0220-0], https://hal.inria.fr/hal-01419551.

[34] M. LABBÉ, D. PONCE LOPEZ, J. PUERTO.*A comparative study of formulations and solution methods for the discrete ordered p-median problem*, in "Computers & operations research", 2016, vol. 78, p. 230–242, https://hal.inria.fr/hal-01419555.

[35] D. PAPADIMITRIOU, B. FORTZ.*Distributed monitoring problem*, in "Electronic Notes in Discrete Mathematics", 2016, vol. 52, p. 13–20, https://hal.inria.fr/hal-01410556.

Articles in Non Peer-Reviewed Journal

[36] S. AFSAR, L. BROTCORNE, P. MARCOTTE, G. SAVARD.*Achieving an optimal trade-off between revenue and energy peak within a smart grid environment*, in "Renewable Energy", March 2016, https://hal.inria.fr/hal-01230915.

Invited Conferences

[37] F. SEMET, M.-I. RESTREPO, T. POCREAU.*Integrated Shift Scheduling and Load Assignment Optimization for Attended Home Delivery*, in "GO X - International Colloquium on Graphs and Optimization 2016", Mount RIgi, Switzerland, July 2016, https://hal.inria.fr/hal-01367274.

[38] F. SEMET.*Challenging optimization problems in E-commerce logistics*, in "46th Annual Conference of the Italian Operations Research Society", Trieste, Italy, Italian Operations Research Society, September 2016, https://hal.inria.fr/hal-01367276.

International Conferences with Proceedings

[39] V. DAL SASSO, L. DE GIOVANNI, M. LABBÉ.*A Column Generation Approach for Pure Parsimony Haplotyping*, in "OASIcs-OpenAccess Series in Informatics", 2016, vol. 50, https://hal.inria.fr/hal-01419553.

[40] B. FORTZ, M. JOYCE-MONIZ, L. GOUVEIA.*On the convex piecewise linear unsplittable multicommodity flow problem*, in "2016 12th International Conference on the Design of Reliable Communication Networks (DRCN)", 2016, p. 9–13, https://hal.inria.fr/hal-01410555.

Conferences without Proceedings

[41] N. ABSI, D. CATTARUZZA, D. FEILLET, M. OGIER, F. SEMET.*Inventory Routing Problem over the long term: a math-heuristic approach*, in "Verolog 2016", Nantes, France, June 2016, https://hal.inria.fr/hal-01335601.

[42] D. C. CATTARUZZA, L. BROTCORNE, N. MITTON, T. RAZAFINDRALAMBO, F. SEMET.*A two-phase matheuristic for the multi-robot routing problem under connectivity constraints*, in "Congrès annuel de la société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF)", Compiègne, France, February 2016, https://hal.inria.fr/hal-01256730.

[43] M.-I. RESTREPO, F. SEMET, T. POCREAU.*Location-allocation and load assignment problem for a package delivery company*, in "Verolog 2016", Nantes, France, June 2016, https://hal.inria.fr/hal-01335612.

[44] F. SEMET, L. BROTCORNE, D. C. CATTARUZZA, B. TOUNSI.*An integrated multi-period stochastic problem in e-fullfilment optimization* , in "TRISTAN IX- Triennial Symposium on Transportation Analysis", Oranjestad, Aruba, June 2016, https://hal.inria.fr/hal-01335627.

Other Publications

[45] S. AFSAR, L. BROTCORNE, P. MARCOTTE, G. SAVARD.*Exact and Heuristic Methods for Energy Pricing Problem*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01230919.

[46] C. CASORRÁN, B. FORTZ, M. LABBÉ, F. ORDÓÑEZ.*Novel formulations for general and security Stackelberg games*, 2016, working paper or preprint, https://hal.inria.fr/hal-01429265.

Project-Team LINKS

Linking Dynamic Data

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
CNRS

Université Charles de Gaulle (Lille 3)

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Data and Knowledge Representation and Processing

Table of contents

Project-Team LINKS

Creation of the Team: 2013 January 01, updated into Project-Team: 2016 June 01

Keywords:

Computer Science and Digital Science:

2.1. - Programming Languages
2.1.1. - Semantics of programming languages
2.1.3. - Functional programming
2.1.6. - Concurrent programming
2.4. - Verification, reliability, certification
2.4.1. - Analysis
2.4.2. - Model-checking
2.4.3. - Proofs
3.1. - Data
3.1.1. - Modeling, representation
3.1.2. - Data management, quering and storage
3.1.3. - Distributed data
3.1.4. - Uncertain data
3.1.5. - Control access, privacy
3.1.6. - Query optimization
3.1.7. - Open data
3.1.8. - Big data (production, storage, transfer)
3.1.9. - Database
3.2.1. - Knowledge bases
3.2.2. - Knowledge extraction, cleaning
3.2.3. - Inference
3.2.4. - Semantic Web
7. - Fundamental Algorithmics
7.4. - Logic in Computer Science
8. - Artificial intelligence
8.1. - Knowledge
8.2. - Machine learning

Other Research Topics and Application Domains:

6.1. - Software industry
6.3.1. - Web
6.3.4. - Social Networks
6.5. - Information systems
9.4.1. - Computer science
9.4.5. - Data science
9.8. - Privacy

1. Members

Research Scientists

Joachim Niehren [Team leader, Inria, Researcher, HDR]
Pierre Bourhis [CNRS, Researcher]

Faculty Members

Iovka Boneva [Univ. Lille I, Associate Professor]
Aurélien Lemay [Univ. Lille III, Associate Professor]
Sylvain Salvati [Univ. Lille I, Professor, HDR]
Slawomir Staworko [Univ. Lille III, Associate Professor, HDR]
Sophie Tison [Univ. Lille I, Professor, HDR]

PhD Students

Adrien Boiret [Univ. Lille I]
Dimitri Gallois [Univ. Lille I]
José Martin Lozano [Univ. Lille I, from Oct 2016]
Momar Sakho [Inria, from Feb 2016]
Tom Sebastian [Univ. Lille I]

Post-Doctoral Fellows

Vincent Hugot [Univ. Lille I, Research Scientist]
Nicolas Bacquey [Inria, from Sep 2016]

Visiting Scientist

Domagoj Vrgoc [from Aug 2016 until Sep 2016]

Administrative Assistant

Aurore Hermant [Inria]

2. Overall Objectives

2.1. Overall Objectives

We will develop algorithms for answering logical querying on heterogeneous linked data collections in hybrid formats, distributed programming languages for managing dynamic linked data collections and workflows based on queries and mappings, and symbolic machine learning algorithms that can link datasets by inferring appropriate queries and mappings.

2.2. Presentation

The following three paragraphs summarise our main research objectives.

Querying Heterogeneous Linked Data We will develop new kinds of schema mappings for semi-structured datasets in hybrid formats including graph databases, RDF collections, and relational databases. These induce recursive queries on linked data collections for which we will investigate evaluation algorithms, containment problems, and concrete applications.

Managing Dynamic Linked Data In order to manage dynamic linked data collections and workflows, we will develop distributed data-centric programming languages with streams and parallelism, based on novel algorithms for incremental query answering, study the propagation of updates of dynamic data through schema mappings, and investigate static analysis methods for linked data workflows.

Linking Data Graphs Finally, we will develop symbolic machine learning algorithms, for inferring queries and mappings between linked data collections in various graphs formats from annotated examples.

3. Research Program

3.1. Background

The main objective of LINKS is to develop methods for querying and managing linked data collections. Even though open linked data is the most prominent example, we will focus on hybrid linked data collections, which are collections of semi-structured datasets in hybrid formats: graph-based, RDF, relational, and NoSQL. The elements of these datasets may be linked, either by pointers or by additional relations between the elements of the different datasets, for instance the “same-as” or “member-of” relations as in RDF.

The advantage of traditional data models is that there exist powerful querying methods and technologies that one might want to preserve. In particular, they come with powerful schemas that constraint the possible manners in which knowledge is represented to a finite number of patterns. The exhaustiveness of these patterns is essential for writing of queries that cover all possible cases. Pattern violations are excluded by schema validation. In contrast, RDF schema languages such as RDFS can only enrich the relations of a dataset by new relations, which also helps for query writing, but which cannot constraint the number of possible patterns, so that they do not come with any reasonable notion of schema validation.

The main weakness of traditional formats, however, is that they do not scale to large data collections as stored on the Web, while the RDF data models scales well to very big collections such as linked open data. Therefore, our objective is to study mixed data collections, some of which may be in RDF format, in which we can lift the advantages of smaller datasets in traditional formats to much larger linked data collections. Such data collections are typically distributed over the internet, that some data sources have rigid query facilities that cannot be easily adapted or extended.

The main assumption that we impose in order to enable the logical approach, is that the given linked data collection must be correct in most dimensions. This means that all datasets are well-formed with respect to their available constraints and schemas, and clean with respect to the data values in most of the components of the relations in the datasets. One of the challenges is to integrate good quality RDF datasets into this setting, another is to clean the incorrect data in those dimensions that are less proper. It remains to be investigated in how far these assumptions can be maintained in realistic applications, and how much they can be weakened otherwise.

For querying linked data collections, the main problems are to resolve the heterogeneity of data formats and schemas, to understand the efficiency and expressiveness of recursive queries, that can follow links repeatedly, to answer queries under constraints, and to optimize query answering algorithms based on static analysis. When linked data is dynamically created, exchanged, or updated, the problems are how to process linked data incrementally, and how to manage linked data collections that change dynamically. In any case (static and dynamic) one needs to find appropriate schema mappings for linking semi-structured datasets. We will study how to automatize parts of this search process by developing symbolic machine learning techniques for linked data collections.

3.2. Querying Heterogeneous Linked Data

Our main objective is to query collections of linked datasets. In the static setting, we consider two kinds of links: explicit links between elements of the datasets, such as equalities or pointers, and logical links between relations of different datasets such as schema mappings. In the dynamic setting, we permit a third kind of links that point to “intentional” relations computable from a description, such as the application of a Web service or the application of a schema mapping.

We believe that collections of linked datasets are usually too big to ensure a global knowledge of all datasets. Therefore, schema mappings and constraints should remain between pairs of datasets. Our main goal is to be able to pose a query on a collection of datasets, while accounting for the possible recursive effects of schema mappings. For illustration, consider a ring of datasets D_1, D_2, D_3 linked by schema mappings M_1, M_2, M_3 that tell us how to complete a database D_i by new elements from the next database in the cycle.

The mappings M_i induce three intentional datasets I_1, I_2, and I_3, such that I_i contains all elements from D_i and all elements implied by M_i from the next intentional dataset in the ring:

$$I_1 = D_1 \cup M_1(I_2), \quad I_2 = D_2 \cup M_2(I_3), \quad I_3 = D_3 \cup M_3(I_1)$$

Clearly, the global information collected by the intentional datasets depends recursively on all three original datasets D_i. Queries to the global information can now be specified as standard queries to the intentional databases I_i. However, we will never materialize the intentional databases I_i. Instead, we can rewrite queries on one of the intentional datasets I_i to recursive queries on the union of the original datasets D_1, D_2, and D_3 with their links and relations. Therefore, a query answering algorithm is needed for recursive queries, that chases the "links" between the D_i in order to compute the part of I_i needed for the purpose of query answering.

This illustrates that we must account for the graph data models when dealing with linked data collections whose elements are linked, and that query languages for such graphs must provide recursion in order to chase links. Therefore, we will have to study graph databases with recursive queries, such as RDF graphs with SPARQL queries, but also other classes of graph databases and queries.

We study schemas and mappings between datasets with different kinds of data models and the complexity of evaluating recursive queries over graphs. In order to use schema mapping for efficiently querying the different datasets, we need to optimize the queries by taking into account the mappings. Therefore, we will study static analysis of schema mappings and recursive queries. Finally, we develop concrete applications in which our fundamental techniques can be applied.

3.3. Managing Dynamic Linked Data

With the quick growth of the information technology on the Web, more and more Web data gets created dynamically every day, for instance by smartphones, industrial machines, users of social networks, and all kinds of sensors. Therefore, large amounts of dynamic data need to be exchanged and managed by various data-centric web services, such as online shops, online newspapers, and social networks.

Dynamic data is often created by the application of some kind of service on the Web. This kind of data is intentional in the same spirit as the intentional data specified by the application of a schema mapping, or the application of some query to the hidden Web. Therefore, we will consider a third kind of links in the dynamic setting, that map to intentional data specified by whatever kind of function application. Such a function can be defined in data-centric programming languages, in the style of Active XML, XSLT, and NOSQL languages.

The dynamicity of data adds a further dimension to the challenges for linked data collections that we described before, while all the difficulties remain valid. One of the new aspects is that intentional data may be produced incrementally, as for instance when exchanged over data streams. Therefore, one needs incremental algorithms able to evaluate queries on incomplete linked data collections, that are extended or updated incrementally. Note that incremental data may be produced without end, such as a Twitter stream, so that one cannot wait for its completion. Instead, one needs to query and manage dynamic data with as low latency as possible. Furthermore, all static analysis problems are to be re-investigated in the presence of dynamic data.

Another aspect of dynamic data is distribution over the Web, and thus parallel processing as in the cloud. This raises the typical problems coming with data distribution: huge data sources cannot be moved without very high costs, while data must be replicated for providing efficient parallel access. This makes it difficult, if not impossible, to update replicated data consistently. Therefore, the consistency assumption has been removed by NOSQL databases for instance, while parallel algorithmic is limited to naive parallelisation (i.e. map/reduce) where only few data needs to be exchanged.

We will investigate incremental query evaluation for distributed data-centered programming languages for linked data collections, dynamic updates as needed for linked data management, and static analysis for linked data workflows.

3.4. Linking Graphs

When datasets from independent sources are not linked with existing schema mappings, we would like to investigate symbolic machine learning solutions for inferring such mappings in order to define meaningful links between data from separate sources. This problem can be studied for various kinds of linked data collections. Before presenting the precise objectives, we will illustrate our approach on the example of linking data in two independent graphs: an address book of a research institute containing detailed personnel information and a (global) bibliographic database containing information on papers and their authors.

We remind that a schema allows to identify a collection of types each grouping objects from the same semantic class e.g., the collection of all persons in the address book and the collection of all authors in the bibliography database. As a schema is often lacking or underspecified in graph data models, we intend to investigate inference methods based on structural similarity of graph fragments used to describe objects from the same class in a given document e.g., in the bibliographic database every author has a name and a number of affiliations, while a paper has a title and a number of authors. Furthermore, our inference methods will attempt to identify, for every type, a set of possible keys, where by key we understand a collection of attributes of an object that uniquely identifies such an object in its semantic class. For instance, for a person in the address book two examples of a key are the name of the person and the office phone number of that person.

In the next step, we plan to investigate employing existing entity linkage solutions to identify pairs of types from different databases whose instances should be linked using compatible keys. For instance, persons in the address book should be linked with authors in the bibliographical database using the name as the compatible key. Linking the same objects (represented in different ways) in two databases can be viewed as an instance of a mapping between the two databases. Such mapping is, however, discriminatory because it typically maps objects from a specific subset of objects of given types. For instance, the mapping implied by linking persons in the address book with authors in the bibliographic database involves in fact researchers, a subgroup of personnel of the research institute, and authors affiliated with the research institute. Naturally, a subset of objects of a given type, or a subtype, can be viewed as a result of a query on the set of all objects, which on very basic level illustrates how learning data mappings can be reduced to learning queries.

While basic mappings link objects of the same type, more general mappings define how the same type of information is represented in two different databases. For instance, the email address and the postal address of an individual may be represented in one way in the address book and in another way in the bibliographic databases, and naturally, the query asking for the email address and the postal address of a person identified by a given name will differ from one database to the other. While queries used in the context of linking objects of compatible types are essentially unary, queries used in the context of linking information are n-ary and we plan to approach inference of general database mappings by investigating and employing algorithms for inference of n-ary queries.

An important goal in this research is elaborating a formal definition of *learnability* (feasibility of inference) of a given class of concepts (schemas of queries). We plan to following the example of Gold (1967), which requires not only the existence of an efficient algorithm that infers concepts consistent with the given input but the ability to infer every concept from the given class with a sufficiently informative input. Naturally, learnability depends on two parameters. The first parameter is the class of concepts i.e., a class of schema and a class of queries, from which the goal concept is to be inferred. The second parameter is the type of input that an inference algorithm is given. This can be a set of examples of a concept e.g., instances of RDF databases for which we wish to construct a schema or a selection of nodes that a goal query is to select. Alternatively, a more general interactive scenario can be used where the learning algorithm inquires the user about the goal concept e.g., by asking to indicate whether a given node is to be selected or not (as membership queries of Angluin (1987)). In general, the richer the input is, the richer class of concepts can be handled, however, the richer class of queries is to be handled, the higher computational cost is to be expected. The primary task is to find a good compromise and identify classes of concepts that are of high practical value, allow efficient inference with possibly simple type of input.

The main open problem for graph-shaped data studied by Links are how to infer queries, schemas, and schema-mappings for graph-structured data.

4. Application Domains

4.1. Linked Data Integration

There are many contexts in which integrating linked data is interesting. We advocate here one possible scenario, namely that of integrating business linked data to feed what is called Business Intelligence. The latter consists of a set of theories and methodologies that transform raw data into meaningful and useful information for business purposes (from Wikipedia). In the past decade, most of the enterprise data was proprietary, thus residing within the enterprise repository, along with the knowledge derived from that data. Today's' enterprises and businessmen need to face the problem of information explosion, due to the Internet's ability to rapidly convey large amounts of information throughout the world via end-user applications and tools. Although linked data collections exist by bridging the gap between enterprise data and external resources, they are not sufficient to support the various tasks of Business Intelligence. To make a concrete example, concepts in an enterprise repository need to be matched with concepts in Wikipedia and this can be done via pointers or equalities. However, more complex logical statements (i.e. mappings) need to be conceived to map a portion of a local database to a portion of an RDF graph, such as a subgraph in Wikipedia or in a social network, e.g. LinkedIn. Such mappings would then enrich the amount of knowledge shared within the enterprise and let more complex queries be evaluated. As an example, businessmen with the aid of business intelligence tools need to make complex sentimental analysis on the potential clients and for such a reason, such tools must be able to pose complex queries, that exploit the previous logical mappings to guide their analysis. Moreover, the external resources may be rapidly evolving thus leading to revisit the current state of business intelligence within the enterprise.

4.2. Data Cleaning

The second example of application of our proposal concerns scientists who want to quickly inspect relevant literature and datasets. In such a case, local knowledge that comes from a local repository of publications belonging to a research institute (e.g. HAL) need to be integrated with other Web-based repositories, such as DBLP, Google Scholar, ResearchGate and even Wikipedia. Indeed, the local repository may be incomplete or contain semantic ambiguities, such as mistaken or missing conference venues, mistaken long names for the publication venues and journals, missing explanation of research keywords, and opaque keywords. We envision a publication management system that exploits both links between database elements, namely pointers to external resources and logical links. The latter can be complex relationships between local portions of data and remote resources, encoded as schema mappings. There are different tasks that such a scenario could entail such as (i) cleaning the errors with links to correct data e.g. via mappings from HAL to DBLP for the publications errors, and via mappings from HAL to Wikipedia for opaque keywords, (ii) thoroughly enrich the list of publications of a given research institute, and (iii) support complex queries on the corrected data combined with logical mappings.

4.3. Real Time Complex Event Processing

Complex event processing serves for monitoring nested word streams in real time. Complex event streams are gaining popularity with social networks such as with Facebook and Twitter, and thus should be supported by distributed databases on the Web. Since this is not yet the case, there remains much space for future industrial transfer related to Links' second axis on dynamic linked data.

5. Highlights of the Year

5.1. Highlights of the Year

Certain Query Answering as Access Control

P. Bourhis [24] presented at **LICS** — the top conference in logic in computer science — a general framework for querying databases with visible and invisible relations. This work was done in cooperation with Oxford, Santa Cruez, and Bordeaux. It generalizes in a uniform manner the problems of certain query answering and access control for relational databases. Invisible relations are subject to the open world assumption possibly under constraints, while visible relations are subject to the closed world assumption. Bourhis then shows that the problem of answering Boolean conjunctive queries in this framework is decidable, and studies the complexity of various versions of this problem. It turns out that the complexity increases compared to the problem of certain query answering, given that the closed world assumption is adopted for the added visible relations.

Five ANR Projects

Two new ANR projects were accepted this year: *Delta* and *Headwork*. This makes Links a partner of 5 ANR projects in 2016.

PhD Defense of A. Boiret

The defense of the PhD thesis of A. Boiret [11] on "Normalization and Learning of Transducers on Trees and Words" under the supervision of J. Niehren and A. Lemay was highly appreciated by the reviewers. In particular, he illustrated very clearly how to learn top-down tree transformations subject to regular schema restriction [31], [33], [34]. Furthermore, he solve a problem open for more than 20 years on how to learn rational functions, i.e. word transducers with regular lookahead.

6. New Software and Platforms

6.1. ShEx Validator

KEYWORDS: RDF Data management - RDF - Shape Expression
FUNCTIONAL DESCRIPTION

Shape Expression schemas is a formalism for defining constraints on RDF graphs. This software allows to check whether a graph satisfies a Shape Expressions schema.

- Participants: Iovka Boneva
- Contact: Iovka Boneva
- URL: https://gforge.inria.fr/projects/shex-impl

6.2. gMark

KEYWORDS: graph benchmark - Graph Database - Graph Query
FUNCTIONAL DESCRIPTION

gMark allow the generation of graph databases and an associated set of query from a schema of the graph. gMark is based on the following principles: great flexibility in the schema definition, ability to generate big size graphs, ability to generate recursive queries and queries with a desired selectivity .

- Participants: Aurélien Lemay
- Contact: Aurélien Lemay
- URL: https://github.com/graphMark/gmark

6.3. QuiXPath

KEYWORDS: XML Streams - XPath 3.0 Queries - Aggregation - Data Joins
FUNCTIONAL DESCRIPTION

QuiXPath is a streaming implementation that covers most of XPath 3.0. It was developed during the PhD thesis of T. Sebastian funded by our industrial transfer partner Innovimax.

- Participants: Tom Sebastian and Joachim Niehren
- Contact: Joachim Niehren
- URL: https://project.inria.fr/quix-tool-suite

6.4. X-FUN

KEYWORDS: XML - Transformation - Functional programming - Compilers - Programming language
FUNCTIONAL DESCRIPTION

X-FUN is a core language for implementing various XML, standards in a uniform manner. X-Fun is a higher-order functional programming language for transforming data trees based on node selection queries.

- Participants: Pavel Labath and Joachim Niehren
- Contact: Joachim Niehren

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Provenance

The computation of the provenance of a query answer is a classical problem in database theory. It consists in aggregating the impact of tuples of a database to a query answer. This allows to give an explanation of the query answers, that can help to judge their reliability. The computation of the provenance of a query answer is thus an aggregation problem as studied by the ANR project *Aggreg* .

P. Bourhis [20] showed at **PODS** — the top conference on database theory — that the lineage of MSO queries on treelike database instances is tractable, but not on other instances. This work was in cooperation with Telecom ParisTech and ENS Paris. As a first application, he can show that MSO query evaluation on probabilistic databases is tractable for tree like database instances, but not otherwise.

P. Bourhis applied in cooperation with Tel Aviv, provenance problems to recommendation systems. This allows to explain the end result by summarising with similar data without changing significantly results obtained in general by aggregation on the data. The corresponding tool was demonstrated at **EDBT** [32].

7.1.2. Certain Query Answering and Access Control

The problem of certain query answering consists in finding which are the certain answer of a query in a database with incomplete data, and a set of constraints representing available the knowledge on the incomplete data.

P. Bourhis [24] presented at **LICS** — the top conference in logic in computer science — a general framework for querying databases with visible and invisible relations. This work was done in cooperation with Oxford, Santa Cruez, and Bordeaux. His framework is motivated by the problem of access control for relational databases, i.e. of data leakage in relational views, but generalizes at the same time the problem of certain query answering. Invisible relations are subject to the open world assumption possibly under constraints as usual in certain query answering, while visible relations are subject to the closed world assumption. Bourhis then show that it is decidable, whether a conjunctive has an answer in this framework, when given the visible relation, the constraints, and the query as inputs. He also studies the complexity of this problem. It turns out the complexity increases from polynomial to doubly exponential, compared to certain query answering, since adding visible relations subject to the closed world assumption.

P. Bourhis studied at **IJCAI** [19] certain query answering with some transitive closure constraints, which allow to define a constraints with recursion. This work was done in collaboration with Oxford and Telecom ParisTech.

The problem of ontological query containment consists in establishing whether the certain answers of two queries subject to an ontology are included in each other. P. Bourhis [26] studied at **KR** this problem for several closely related formalisms: monadic disjunctive Datalog (MDDLog), MMSNP (a logical generalization of constraint satisfaction problems) and ontology-mediated queries (OMQs). This work was done in cooperation with Bremen.

7.1.3. Recursive Queries

At **LICS** [21] again, P. Bourhis showed in collaboration with Oxford how to lift a major restriction on decidable fixpoint logics that can define recursive queries (such as C2RPQs), specifically on guarded logic. This allows to improve significantly expressiveness of decidable fixpoint logics.

A. Lemay contributed at **TKDE** [14] the *gMark* benchmark, a tool to generate large size graph database and an associated set of queries. This work was done in cooperation with Eindhoven and previous members of Links that are now in Lyon and Clérmont-Ferrant. The tool was also demonstrated at **VLDB** [13]. Its main interest is a great flexibility (the generation of the graph can be done from a simple schema, but can also incorporate elaborate a parameters), an ability to generate recursive queries, and the possibility to generate large sets of queries of a desired selectivity. This benchmark allowed for instance to highlight difficulties for the existing query engines to deal with recursive queries of high selectivity.

7.1.4. Data Integration

P. Bourhis and S. Staworko in cooperation with Bordeaux and Oxford presented at **TODS** [17] their work on bounded repairability for regular tree languages, which is a study on whether a tree document (typically XML) can be repaired to fit a given target tree language within a bounded amount of tree editing operations. The article studies the complexity of different classes of tree languages such as non-recursive DTDs, recursive DTDs, or languages by arbitrary bottom-up tree automaton.

J.M. Lozano started his PhD project under the supervision of I. Boneva and S. Staworko. His topic subscribes the ANR project *Datacert* on data integration and certification.

7.1.5. Schema Validation

A. Boiret, V. Hugot and J. Niehren studied schemas for JSON documents in **Information and Computation** [15]. This work was done in collaboration with Paris 7. A JSON document is an unordered data trees, so schemas for such documents are best seen as automata for unordered data trees. The paper generalizes several previous formalisms for automata on unordered trees in a uniform framework. Whether the equivalence of two schemas can be tested in P-time is studied for various instances of the framework.

This work subscribes to the ANR project *Colis* where unranked data trees are used as models of linux file systems. In this context, N. Bacquey started his postdoc on the verification of linux installation scripts.

7.2. Managing Dynamic Linked Data

7.2.1. Complex Event Processing

Complex event processing can be seen as the problem is to answer queries on data graphs, for graphs that arrive on streams. These queries may contain aggregates, so this work subscribes to the ANR project *Aggreg*.

In his PhD thesis, T. Sebastian [12] developed with his supervisor J. Niehren streaming algorithms covering all of XPath 3.0 queries on `XML` streams. For this, they proposed a higher-order query language λXP, showed how to give a formal semantics of all of XPath 3.0 by compilation to λXP, and then how to evaluate λXP queries on `XML` streams. These algorithms were implemented in the QuiXPath tool.

At **SOFSEM**, they proposed a new technique to speed up the evaluation of navigational XPath queries on XML streams based on document projection. The idea is to skip those parts of the stream that are irrelevant for the query. This speeds up the evaluation of navigation XPath queries by a factor of 4 in usual Xpath benchmarks.

M. Sakho started his PhD project on hyperstreaming query answering algorithms for graphs under the supervision of J. Niehren and I. Boneva. Part of this work will be continued with out visitor D. Vrgoc from Santiago di Chili.

7.2.2. *Data Centric Workflows*

Data-centric workflows are complex programs that can query and update a database. The usage of data-centric workflows for crowd sourcing is the topic of the ANR Project *HeadWork*.

In collaboration with ENS Cachan and San Diego, P. Bourhis presented at **ICDT** [18] techniques on collaborative access control in a distributed query and data exchange language (Webdamlog). The goal of this work was to provide a semantic to data exchange rules defined by Webdamlog. It also allowed to prove that it is possible to formally verify whether there are data leakages.

P. Bourhis with Tel Aviv defined at **ICDE** [25] a notion of provenance for data-centric workflows, and proved that it can be used to explain the provenance of fact in the final instance of an execution. This provenance is used to answer three main questions: *why* does a specific tuple appear in the answer of a query, *what if* the initial database is changed (Revision problem), and *how to* change the query to obtain a missing tuple.

7.3. Linking Data Graphs

7.3.1. *Learning Transformations*

We consider the problem to learn queries and query-based transformations on semi-structured data from examples.

A. Boiret obtained his PhD for his work on the "Normalization and Learning of Transducers on Trees and Words" under the supervision of J. Niehren and A. Lemay. In this year, he showed how to learn top-down tree transformations with regular schema restrictions [31], [33], [34]. At **LATA** [22], he deepened a result of a previous PhD student of Links on learning sequential tree-to-word transducers (with output concatenation), by showing who to find normal forms for less restrictive linear tree-to-word transducers. At **DLT** [23], he could show in cooperation with Munich, that the equivalence problem of this class of transducers is in polynomial time, even though their normal forms may be of exponential size.

In the context of learning RDF graph transformations, S. Staworko presented a cooperation with Edinburg at **VLDB** [27]. Using bisimulation technique, he aims at aligning datas of two RDF Graphs that takes into account blank velues, changes in ontology and small differences in data values and in the structure of the graph. the alignement of graphs is an important first step for the inference of transformations.

7.3.2. *Learning Join Queries*

S. Staworko published in **TODS** an article [16] on learning join queries from user examples in collaboration with Universities of Lyon and Clermont-Ferrand that present techniques that allow the automatic construction of a join query through interaction with a user that simply labels sets of tuples to indicate whether the tuple is in the target query or not.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Links participates in the CPER DATA (2015-19)

8.2. National Initiatives

ANR Aggreg (2014-19): Aggregated Queries.

- Particpants: J. Niehren [correspondent], P. Bourhis, A. Lemay, A. Boiret
- The coordinator is J. Niehren and theparteners are the University Paris 7 (A. Durand) including members of the Inria project DAHU (L. Ségoufin), the University of Marseille (N. Creignou) and University of Caen (E. Grandjean).
- Objective: the main goal of the Aggreg project is to develop efficient algorithms and to study the complexity of answering aggregate queries for databases and data streams of various kinds.

ANR Colis (2015-20): Correctness of Linux Scripts.

- Participants: J. Niehren [correspondent], A. Lemay, S. Tison, A. Boiret, V. Hugot.
- The coordinator is R. Treinen from the University of Paris 7 and the other partner is the Tocata project of Inria Saclay (C. Marché).
- Objective: This project aims at verifying the correctness of transformations on data trees defined by shell scripts for Linux software installation. The data trees here are the instance of the file system which are changed by installation scripts.

ANR DataCert (2015-20):

- Participants: I. Boneva [correspondent], S. Tison, J. Lozano.
- Partners: The coordinator is E. Contejean from the University of Paris Sud and the other partner is the University of Lyon.
- Objective: the main goals of the Datacert project are to provide deep specification in Coq of algorithms for data integration and exchange and of algorithms for enforcing security policies, as well as to design data integration methods for data models beyond the relational data model.

ANR Headwork (2016-21):

- Participants: P. Bourhis [correspondant], J. Niehren, M. Sakho.
- Scientific partners: The coordinateur is D. Gross-Amblard from the Druid Team (Rennes 1). Other partners include the Dahu team (Inria Saclay) and Sumo (Inria Bretagne)
- Industrial partners: Spipoll, and Foulefactory.
- Objective: The main object is to develop data-centric workflows for programming crowd sourcing systems in flexible declarative manner. The problem of crowd sourcing systems is to fill a database with knowledge gathered by thousands or more human participants. A particular focus is to be put on the aspects of data uncertainty and for the representation of user expertise.

ANR Delta (2016-21):

- Participants: P. Bourhis [correspondent], D. Gallois.
- Partners: The coordinator is M. Zeitoun from LaBRI, other partners are LIF (Marseille) and IRIF (Paris-Diderot).
- Objective: Delta is focused on the study of logic, transducers and automata. In particular, it aims at extending classical framework to handle input/output, quantities and data.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Declared Inria International Partners

AMSud project “Foundations of Graph Databases” (2015-16)

Partners: Chili (C. Riveros), Buenos Aires (Figueira), Bordeaux (G. Puppis).

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Domagoj Vrgoc, DCC PUC Chile, From Aug 2016 Until Sep 2016

8.4.2. Visits to International Teams

8.4.2.1. Research Stays Abroad

Slawek Staworko, University of Edinburgh, 2014-16.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. Member of the Conference Program Committees

J. Niehren was member of the program committees of LPAR (International Conference on Logic Programming and Automatic Reasoning) 2016.

S. Tison is member of the program committees of FSCD (First International Conference on Formal Structures for Computation and Deduction) 2016.

S. Staworko is member of the program committes of PODS (ACM Symposium on Principles of Database Systems) 2016.

I. Boneva was member of program committee of EDBT (International Conference on Extending Database Technology) 2016 Vision Track.

P. Bourhis was member of program committee of Provenance week 2016, IJCAI (International Joint Conference on Artificial Intelligence) 2016.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

S. Tison is in the editorial committee of RAIRO-ITA (Theoretical Informatics and Applications).

J. Niehren is in the editoral board of Fundamenta Informaticae.

9.1.2.2. Reviewer - Reviewing Activities

Too many to be enumerated.

9.1.3. Leadership within the Scientific Community

S. Tison has been member of the “Comité National de la Recherche Scientifique (CoNRS)” (Section 6) until June 2016.

9.1.4. Scientific Expertise

P. Bourhis has expertised a project in the ’Recherche Formation Innovation Atlanstic’ program of Pays de la Loire Region.

S. Tison has been a member of the scientific board of the company See-d

9.1.5. Research Administration

S. Tison is an elected member of the academic council of "ComUE Lille Nord de France " since November 2015.

S. Tison is a vice president of the University of Lille 1 since October 2015, where she is responsible for industrial partnerships, innovation, and valorisation.

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Bachelor : S. Tison, Discrete Mathematics, 36h, Université Lille 1, France

Bachelor : S. Salvati, Computer Science Info, L1, 96h, Université de Lille 1, France

Bachelor : S. Salvati, Automata and Languages, L3, 36h, Université de Lille 1, France

Bachelor : S. Salvati, Algorithmic and Operational Research, L3, 36h, Université de Lille 1, France

Master 2 (MOCAD): P. Bourhis and J. Niehren Information extraction, 20h30, M2, Université Lille 1, France

Master 2 (MOCAD) : I. Boneva and P. Bourhis, Algorithms for Database, 21h, M1, Université Lille 1, France

Master 1 : S. Tison, Advanced algorithms and complexity, M1, 54h, Université Lille 1, France

Master 1 : S. Tison, Business Intelligence, M1, 28h, Université Lille 1, France

Master 1 : A. Lemay, XML Technologies, 16h, M2, Université Lille 3, France

Master 1 : S. Staworko is co-head of the master ’Web Analyst’ in Université Lille 3, France

DUT : I. Boneva, 100h, Université Lillle 1, France

A. Lemay is pedagogical responsable for Computer Science and numeric correspondant for UFR LEA Lille 3.

9.2.2. Supervision

PhD defended : T. Sebastian, Streaming algorithms for XPath. 2011-2016. Supervised by Niehren.

PhD defended: A. Boiret. Normalization and Learning of Transducers on Trees and Words. 2011-2016. Supervised by Niehren and Lemay.

PhD in progress: D. Gallois. Since 2015. Recursive Queries. Supervised by Bourhis and Tison.

PhD in progress: M. Sakho. Hyperstreaming Query answering on graphs. Since 2016. Supervised by Niehren and Boneva.

PhD in progress: J.M. Lozano. On data integration for mixed database formats. Supervised by Boneva and Staworko.

9.2.3. Juries

P. Bourhis was a member of PhD committee for Antoine Amarilli (ParisTech)

S. Tison was a member of the PhD committees for Yoann Dufresne and A. Boiret and reviewer for the PhD of Carles Creus (Universitat Politecnica de Catalunya).

A. Lemay was a member of the PhD committee for A. Boiret

J. Niehren was a member of PhD committees for A. Boiret, T. Sebastian and Louis Gallaraga (ParisTech).

9.2.4. Internships

A. Durey, University of Lille I. On Implemetation of a Shex validator, from Jun 2016 until Sep 2016. Supervison I. Boneva.

R. Li, Ecole Centrale de Lille, from May 2016 until Aug 2016. On aggregate queries for data mining. Supervisors J. Niehren and P. Bourhis.

9.2.5. Selection Committies

I. Boneva was member of the selection committee for an assistant professorship at the University of Lille I.

S. Tison was member of the selection committee for a professorship at the University of Rouen.

9.3. Popularization

I. Boneva has supervised second year students from DUT while they conducted activities on introduction to programming to 9-10 years old students in Villeneuve d'Ascq.

9.4. Standardization

I. Boneva is a member of the Data Shapes Working Group of the W3C which the mission is to produce a language for defining structural constraints on RDF graphs. http://www.w3.org/2014/data-shapes/charter

10. Bibliography

Major publications by the team in recent years

[1] A. AMARILLI, P. BOURHIS, P. SENELLART.*Tractable Lineages on Treelike Instances: Limits and Extensions*, in "PODS (Principles of Database Systems)", San Francisco, United States, June 2016, p. 355-370, https://hal-institut-mines-telecom.archives-ouvertes.fr/hal-01336514.

[2] M. BENEDIKT, P. BOURHIS, M. VANDEN BOOM.*A Step Up in Expressiveness of Decidable Fixpoint Logics*, in "Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science", New York, United States, July 2016, https://hal.inria.fr/hal-01413890.

[3] A. BOIRET, V. HUGOT, J. NIEHREN, R. TREINEN.*Automata for Unordered Trees*, in "Information and Computation", July 2016 [*DOI :* 10.1016/J.IC.2016.07.012], https://hal.inria.fr/hal-01179493.

[4] A. BONIFATI, R. CIUCANU, S. STAWORKO.*Learning Join Queries from User Examples*, in "ACM Transactions on Database Systems", February 2016, vol. 40, n^{o} 4, p. 24:1–24:38, https://hal.inria.fr/hal-01187986.

[5] P. BOURHIS, M. BENEDIKT, B. TEN CATE, G. PUPPIS.*Querying Visible and Invisible Information*, in "LICS 2016 - 31st Annual ACM/IEEE Symposium on Logic in Computer Science", New York City, United States, July 2016, p. 297-306 [*DOI :* 10.1145/2933575.2935306], https://hal.archives-ouvertes.fr/hal-01411118.

[6] P. BOURHIS, M. KRÖTZSCH, S. RUDOLPH.*Reasonable Highly Expressive Query Languages*, in "IJCAI", Buenos Aires, Argentina, July 2015, IJCAI-2015 Honorable Mention [*DOI :* 10.1007/978-3-662-47666-6_5], https://hal.inria.fr/hal-01211282.

[7] P. BOURHIS, C. LUTZ.*Containment in Monadic Disjunctive Datalog, MMSNP, and Expressive Description Logics*, in "Principles of Knowledge Representation and Reasoning", Cape Town, South Africa, April 2016, https://hal.inria.fr/hal-01413887.

[8] P. BUNEMAN, S. STAWORKO.*RDF Graph Alignment with Bisimulation*, in "VLDB 2016 - 42nd International Conference on Very Large Databases", New Dehli, India, Proceedings of the VLDB Endowment, September 2016, vol. 9, n^{o} 12, p. 1149 - 1160 [*DOI :* 10.14778/2994509.2994531], https://hal.inria.fr/hal-01417156.

[9] D. DEBARBIEUX, O. GAUWIN, J. NIEHREN, T. SEBASTIAN, M. ZERGAOUI.*Early Nested Word Automata for XPath Query Answering on XML Streams*, in "Theoretical Computer Science", March 2015, n^{o} 578, p. 100-127, https://hal.inria.fr/hal-00966625.

[10] S. STAWORKO, I. BONEVA, J. E. LABRA GAYO, S. HYM, E. G. PRUD'HOMMEAUX, H. SOLBRIG.*Complexity and Expressiveness of ShEx for RDF*, in "18th International Conference on Database Theory (ICDT 2015)", Brussels, Belgium, M. ARENAS, M. UGARTE (editors), 18th International Conference on Database Theory (ICDT 2015), March 2015 [*DOI :* 10.4230/LIPICS.ICDT.2015.195], https://hal.archives-ouvertes.fr/hal-01218552.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. BOIRET.*Normalization and Learning of Transducers on Trees and Words*, Université de Lille, November 2016, https://tel.archives-ouvertes.fr/tel-01396543.

[12] T. SEBASTIAN.*Evaluation of XPath Queries on XML Streams with Networks of Early Nested Word Automata*, Universite Lille 1, June 2016, https://hal.inria.fr/tel-01342511.

Articles in International Peer-Reviewed Journal

[13] G. BAGAN, A. BONIFATI, R. CIUCANU, G. FLETCHER, A. LEMAY, N. ADVOKAAT.*Generating Flexible Workloads for Graph Databases*, in "Proceedings of the VLDB Endowment (PVLDB)", June 2016, https://hal.inria.fr/hal-01330111.

[14] G. BAGAN, A. BONIFATI, R. CIUCANU, G. FLETCHER, A. LEMAY, N. ADVOKAAT.*gMark: Schema-Driven Generation of Graphs and Queries*, in "IEEE Transactions on Knowledge and Data Engineering", November 2016, https://hal.inria.fr/hal-01402575.

[15] A. BOIRET, V. HUGOT, J. NIEHREN, R. TREINEN.*Automata for Unordered Trees*, in "Information and Computation", July 2016 [*DOI :* 10.1016/J.IC.2016.07.012], https://hal.inria.fr/hal-01179493.

[16] A. BONIFATI, R. CIUCANU, S. STAWORKO.*Learning Join Queries from User Examples*, in "ACM Transactions on Database Systems", February 2016, vol. 40, n^{o} 4, p. 24:1–24:38, https://hal.inria.fr/hal-01187986.

[17] P. BOURHIS, C. RIVEROS, S. STAWORKO, G. PUPPIS.*Bounded Repairability for Regular Tree Languages*, in "ACM Transactions on Database Systems", June 2016, vol. 41, n^{o} 3, p. 1-45 [*DOI :* 10.1145/2898995], https://hal.archives-ouvertes.fr/hal-01411116.

International Conferences with Proceedings

[18] S. ABITEBOUL, P. BOURHIS, V. VIANU.*A formal study of collaborative access control in distributed datalog*, in "ICDT 2016 - 19th International Conference on Database Theory", Bordeaux, France, W. MARTENS, T. ZEUME (editors), March 2016, https://hal.inria.fr/hal-01290497.

[19] A. AMARILLI, M. BENEDIKT, P. BOURHIS, M. VANDEN BOOM.*Query Answering with Transitive and Linear-Ordered Data*, in "Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016", New York, United States, July 2016, https://hal.inria.fr/hal-01413881.

[20] A. AMARILLI, P. BOURHIS, P. SENELLART.*Tractable Lineages on Treelike Instances: Limits and Extensions*, in "PODS (Principles of Database Systems)", San Francisco, United States, June 2016, p. 355-370, https://hal-institut-mines-telecom.archives-ouvertes.fr/hal-01336514.

[21] M. BENEDIKT, P. BOURHIS, M. VANDEN BOOM.*A Step Up in Expressiveness of Decidable Fixpoint Logics*, in "Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science", New York, United States, July 2016, https://hal.inria.fr/hal-01413890.

[22] A. BOIRET.*Normal Form on Linear Tree-to-word Transducers*, in "10th International Conference on Language and Automata Theory and Applications", Prague, Czech Republic, J. JANOUŠEK, C. MARTÍN-VIDE (editors), March 2016, https://hal.inria.fr/hal-01218030.

[23] A. BOIRET, R. PALENTA.*Deciding Equivalence of Linear Tree-to-Word Transducers in Polynomial Time*, in "Developments in Language Theory", Montreal, Canada, Lecture Notes in Computer Science, Springer, July 2016, vol. 9840, p. 355-367, https://hal.archives-ouvertes.fr/hal-01429110.

[24] P. BOURHIS, M. BENEDIKT, B. TEN CATE, G. PUPPIS.*Querying Visible and Invisible Information*, in "LICS 2016 - 31st Annual ACM/IEEE Symposium on Logic in Computer Science", New York City, United States, July 2016, p. 297-306 [*DOI :* 10.1145/2933575.2935306], https://hal.archives-ouvertes.fr/hal-01411118.

[25] P. BOURHIS, D. DEUTCH, Y. MOSKOVITCH:.*Analyzing data-centric applications: Why, what-if, and how-to.* , in "32nd IEEE International Conference on Data Engineering, ICDE 2016", Helsinki, Finland, May 2016, https://hal.inria.fr/hal-01413879.

[26] P. BOURHIS, C. LUTZ.*Containment in Monadic Disjunctive Datalog, MMSNP, and Expressive Description Logics*, in "Principles of Knowledge Representation and Reasoning", Cape Town, South Africa, April 2016, https://hal.inria.fr/hal-01413887.

[27] P. BUNEMAN, S. STAWORKO.*RDF Graph Alignment with Bisimulation*, in "VLDB 2016 - 42nd International Conference on Very Large Databases", New Dehli, India, Proceedings of the VLDB Endowment, September 2016, vol. 9, n^{o} 12, p. 1149 - 1160 [*DOI :* 10.14778/2994509.2994531], https://hal.inria.fr/hal-01417156.

[28] T. SEBASTIAN, J. NIEHREN.*Projection for Nested Word Automata Speeds up XPath Evaluation on XML Streams*, in "International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM)", Harrachov, Czech Republic, January 2016, https://hal.inria.fr/hal-01182529.

Conferences without Proceedings

[29] G. Bagan, A. Bonifati, R. Ciucanu, G. Fletcher, A. Lemay, N. Advokaat.*Génération de Requêtes pour les Bases de Données Orientées Graphes*, in "32ème Conférence sur la Gestion de Données - Principes, Technologies et Applications - BDA 2016", Futuroscope, Poitiers, France, November 2016, https://hal.inria.fr/hal-01402582.

[30] G. Bagan, A. Bonifati, R. Ciucanu, G. Fletcher, A. Lemay, N. Advokaat.*gMark : Génération de Graphes et de Requêtes Dirigée par le Schéma*, in "32ème Conférence sur la Gestion de Données - Principes, Technologies et Applications - BDA 2016", Futuroscope, Poitiers, France, November 2016, https://hal.inria.fr/hal-01402580.

[31] A. Boiret, A. Lemay, J. Niehren.*Learning Top-Down Tree Transducers with Regular Domain Inspection*, in "International Conference on Grammatical Inference 2016", Delft, Netherlands, October 2016, https://hal.inria.fr/hal-01357186.

Other Publications

[32] E. Ainy, P. Bourhis, S. B. Davidson, D. Deutch, T. Milo.*PROX: Approximated Summarization of Data Provenance*, March 2016, International Conference on Extending Database Technology, Poster - Démonstration, https://hal.inria.fr/hal-01420452.

[33] A. Boiret, A. Lemay, J. Niehren.*A Learning Algorithm for Top-Down Tree Transducers*, August 2016, working paper or preprint, https://hal.inria.fr/hal-01357627.

[34] A. Boiret, A. Lemay, J. Niehren.*Learning Top-Down Tree Transformations with Regular Inspection*, August 2016, working paper or preprint, https://hal.inria.fr/hal-01357631.

[35] I. Boneva.*Comparative expressiveness of ShEx and SHACL (Early working draft)*, March 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01288285.

Project-Team MAGNET

Machine Learning in Information Networks

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
CNRS

Université Charles de Gaulle (Lille 3)

RESEARCH CENTER
Lille - Nord Europe

THEME
Data and Knowledge Representation and Processing

Table of contents

Project-Team MAGNET

Creation of the Team: 2013 January 01, updated into Project-Team: 2016 May 01

Keywords:

Computer Science and Digital Science:

1.2.9. - Social Networks
3. - Data and knowledge
3.1. - Data
3.1.3. - Distributed data
3.1.4. - Uncertain data
3.2.3. - Inference
3.2.4. - Semantic Web
3.3. - Data and knowledge analysis
3.3.1. - On-line analytical processing
3.3.3. - Big data analysis
3.4. - Machine learning and statistics
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.4. - Optimization and learning
3.5. - Social networks
3.5.1. - Analysis of large graphs
3.5.2. - Recommendation systems
4.8. - Privacy-enhancing technologies
5.8. - Natural language processing
6.2.6. - Optimization
6.3.1. - Inverse problems
7. - Fundamental Algorithmics
7.2. - Discrete mathematics, combinatorics
7.8. - Information theory
7.9. - Graph theory
7.10. - Network science
7.11. - Performance evaluation
8.1. - Knowledge
8.2. - Machine learning
8.4. - Natural language processing
8.6. - Decision support

Other Research Topics and Application Domains:

1. - Life sciences
1.1.11. - Systems biology
2. - Health
2.2.4. - Infectious diseases, Virology
2.3. - Epidemiology

2.4.1. - Pharmaco kinetics and dynamics
2.4.2. - Drug resistance
5.8. - Learning and training
5.10. - Biotechnology
6.3. - Network functions
7.1.2. - Road traffic
8.3. - Urbanism and urban planning
9.4.1. - Computer science
9.4.4. - Chemistry
9.5.8. - Linguistics
9.5.10. - Digital humanities
9.8. - Privacy

1. Members

Research Scientists
Aurelien Bellet [Inria, Researcher]
Pascal Denis [Inria, Researcher]
Jan Ramon [Inria, Senior Researcher]

Faculty Members
Marc Tommasi [Team leader, Univ. Lille III, Professor, HDR]
Remi Gilleron [Univ. Lille III, Professor, HDR]
Mikaela Keller [Univ. Lille III, Associate Professor]
Fabien Torre [Univ. Lille III, Associate Professor]
Fabio Vitale [Univ. Lille III, Associate Professor]

PhD Students
David Chatel [Univ. Lille I]
Mathieu Dehouck [Univ. Lille I]
Geraud Le Falher [Inria]
Thibault Lietard [Normale Sup Rennes, from Feb 2016]
Pauline Wauquier [Clic and Walk]

Visiting Scientists
Soravit Changpinyo [USC, Oct 2016]
Wilhelmiina Hamalainen [University of Helsinki, Nov 2016]

Administrative Assistant
Julie Jonas [Inria]

Others
Pierre Dellenbach [Univ. Lille III, student intern, from Mar 2016 until Aug 2016]
Paul Vanhaesebrouck [Univ. Lille III, student intern, from Mar 2016 until Jul 2016]

2. Overall Objectives

2.1. Presentation

MAGNET is a research group that aims to design new machine learning based methods geared towards mining information networks. Information networks are large collections of interconnected data and documents like citation networks and blog networks among others. Our goal is to propose new prediction methods for texts and networks of texts based on machine learning algorithms in graphs. Such algorithms include node and link classification, link prediction, clustering and probabilistic modeling of graphs. We aim to tackle real-world problems such as browsing, monitoring and recommender systems, and more broadly information extraction in information networks. Application domains cover natural language processing, social networks for cultural data and e-commerce, and biomedical informatics.

3. Research Program

3.1. Introduction

The main objective of MAGNET is to develop original machine learning methods for networked data in order to build applications like browsing, monitoring and recommender systems, and more broadly information extraction in information networks. We consider information networks in which the data consist of both feature vectors and texts. We model such networks as (multiple) (hyper)graphs wherein nodes correspond to entities (documents, spans of text, users, ...) and edges correspond to relations between entities (similarity, answer, co-authoring, friendship, ...). Our main research goal is to propose new on-line and batch learning algorithms for various problems (node classification / clustering, link classification / prediction) which exploit the relationships between data entities and, overall, the graph topology. We are also interested in searching for the best hidden graph structure to be generated for solving a given learning task. Our research will be based on generative models for graphs, on machine learning for graphs and on machine learning for texts. The challenges are the dimensionality of the input space, possibly the dimensionality of the output space, the high level of dependencies between the data, the inherent ambiguity of textual data and the limited amount of human labeling. An additional challenge will be to design scalable methods for large information networks. Hence, we will explore how sampling, randomization and active learning can be leveraged to improve the scalability of the proposed algorithms.

Our research program is organized according to the following questions:

1. How to go beyond vectorial classification models in Natural Language Processing (NLP) tasks?
2. How to adaptively build graphs with respect to the given tasks? How to create networks from observations of information diffusion processes?
3. How to design methods able to achieve a good trade-off between predictive accuracy and computational complexity?
4. How to go beyond strict node homophilic/similarity assumptions in graph-based learning methods?

3.2. Beyond Vectorial Models for NLP

One of our overall research objectives is to derive graph-based machine learning algorithms for natural language and text information extraction tasks. This section discusses the motivations behind the use of graph-based ML approaches for these tasks, the main challenges associated with it, as well as some concrete projects. Some of the challenges go beyond NLP problems and will be further developed in the next sections. An interesting aspect of the project is that we anticipate some important cross-fertilizations between NLP and ML graph-based techniques, with NLP not only benefiting from but also pushing ML graph-based approaches into new directions.

Motivations for resorting to graph-based algorithms for texts are at least threefold. First, online texts are organized in networks. With the advent of the web, and the development of forums, blogs, and micro-blogging, and other forms of social media, text productions have become strongly connected. Interestingly, NLP research has been rather slow in coming to terms with this situation, and most of the literature still focus on document-based or sentence-based predictions (wherein inter-document or inter-sentence structure is not exploited). Furthermore, several multi-document tasks exist in NLP (such as multi-document summarization and cross-document coreference resolution), but most existing work typically ignore document boundaries and simply apply a document-based approach, therefore failing to take advantage of the multi-document dimension [40], [43].

A second motivation comes from the fact that most (if not all) NLP problems can be naturally conceived as graph problems. Thus, NLP tasks often involve discovering a relational structure over a set of text spans (words, phrases, clauses, sentences, etc.). Furthermore, the *input* of numerous NLP tasks is also a graph; indeed, most end-to-end NLP systems are conceived as pipelines wherein the output of one processor is in the input of the next. For instance, several tasks take POS tagged sequences or dependency trees as input. But this structured input is often converted to a vectorial form, which inevitably involves a loss of information.

Finally, graph-based representations and learning methods appear to address some core problems faced by NLP, such as the fact that textual data are typically not independent and identically distributed, they often live on a manifold, they involve very high dimensionality, and their annotations is costly and scarce. As such, graph-based methods represent an interesting alternative to, or at least complement, structured prediction methods (such as CRFs or structured SVMs) commonly used within NLP. Graph-based methods, like label propagation, have also been shown to be very effective in semi-supervised settings, and have already given some positive results on a few NLP tasks [22], [45].

Given the above motivations, our first line of research will be to investigate how one can leverage an underlying network structure (e.g., hyperlinks, user links) between documents, or text spans in general, to enhance prediction performance for several NLP tasks. We think that a "network effect", similar to the one that took place in Information Retrieval (with the Page Rank algorithm), could also positively impact NLP research. A few recent papers have already opened the way, for instance in attempting to exploit Twitter follower graph to improve sentiment classification [44].

Part of the challenge here will be to investigate how adequately and efficiently one can model these problems as instances of more general graph-based problems, such as node clustering/classification or link prediction discussed in the next sections. In a few cases, like text classification or sentiment analysis, graph modeling appears to be straightforward: nodes correspond to texts (and potentially users), and edges are given by relationships like hyperlinks, co-authorship, friendship, or thread membership. Unfortunately, modeling NLP problems as networks is not always that obvious. From the one hand, the right level of representation will probably vary depending on the task at hand: the nodes will be sentences, phrases, words, etc. From the other hand, the underlying graph will typically not be given a priori, which in turn raises the question of how we construct it. A preliminary discussion of the issue of optimal graph construction for semi-supervised learning in NLP is given in [22], [48]. We identify the issue of adaptive graph construction as an important scientific challenge for machine learning on graphs in general, and we will discuss it further in Section 3.3.

As noted above, many NLP tasks have been recast as structured prediction problems, allowing to capture (some of the) output dependencies. How to best combine structured output and graph-based ML approaches is another challenge that we intend to address. We will initially investigate this question within a semi-supervised context, concentrating on graph regularization and graph propagation methods. Within such approaches, labels are typically binary or in a small finite set. Our objective is to explore how one propagates an exponential number of *structured labels* (like a sequence of tags or a dependency tree) through graphs. Recent attempts at blending structured output models with graph-based models are investigated in [45], [33]. Another related question that we will address in this context is how does one learn with *partial labels* (like partially specified tag sequence or tree) and use the graph structure to complete the output structure. This last question is very relevant to NLP problems where human annotations are costly; being able to learn from partial annotations could therefore allow for more targeted annotations and in turn reduced costs [35].

The NLP tasks we will mostly focus on are coreference resolution and entity linking, temporal structure prediction, and discourse parsing. These tasks will be envisioned in both document and cross-document settings, although we expect to exploit inter-document links either way. Choices for these particular tasks is guided by the fact that they are still open problems for the NLP community, they potentially have a high impact for industrial applications (like information retrieval, question answering, etc.), and we already have some expertise on these tasks in the team (see for instance [34], [30], [32]). As a midterm goal, we also plan to work on tasks more directly relating to micro-blogging, such sentiment analysis and the automatic thread structuring of technical forums; the latter task is in fact an instance of rhetorical structure prediction [47].

We have already initiated some work on the coreference resolution with graph-based learning, by casting the problem as an instance of spectral clustering [32].

3.3. Adaptive Graph Construction

In most applications, edge weights are computed through a complex data modeling process and convey crucially important information for classifying nodes, making it possible to infer information related to each data sample even exploiting the graph topology solely. In fact, a widespread approach to several classification problems is to represent the data through an undirected weighted graph in which edge weights quantify the similarity between data points. This technique for coding input data has been applied to several domains, including classification of genomic data [42], face recognition [31], and text categorization [36].

In some cases, the full adjacency matrix is generated by employing suitable similarity functions chosen through a deep understanding of the problem structure. For example for the TF-IDF representation of documents, the affinity between pairs of samples is often estimated through the cosine measure or the χ^2 distance. After the generation of the full adjacency matrix, the second phase for obtaining the final graph consists in an edge sparsification/reweighting operation. Some of the edges of the clique obtained in the first step are pruned and the remaining ones can be reweighted to meet the specific requirements of the given classification problem. Constructing a graph with these methods obviously entails various kinds of loss of information. However, in problems like node classification, the use of graphs generated from several datasets can lead to an improvement in accuracy ([49], [23], [24]). Hence, the transformation of a dataset into a graph may, at least in some cases, partially remove various kinds of irregularities present in the original datasets, while keeping some of the most useful information for classifying the data samples. Moreover, it is often possible to accomplish classification tasks on the obtained graph using a running time remarkably lower than is needed by algorithms exploiting the initial datasets, and a suitable sparse graph representation can be seen as a compressed version of the original data. This holds even when input data are provided in a online/stream fashion, so that the resulting graph evolves over time.

In this project we will address the problem of adaptive graph construction towards several directions. The first one is about how to choose the best similarity measure given the objective learning task. This question is related to the question of metric and similarity learning ([25], [26]) which has not been considered in the context of graph-based learning. In the context of structured prediction, we will develop approaches where output structures are organized in graphs whose similarity is given by top-k outcomes of greedy algorithms.

A different way we envision adaptive graph construction is in the context of semi-supervised learning. Partial supervision can take various forms and an interesting and original setting is governed by two currently studied applications: detection of brain anomaly from connectome data and polls recommendation in marketing. Indeed, for these two applications, a partial knowledge of the information diffusion process can be observed while the network is unknown or only partially known. An objective is to construct (or complete) the network structure from some local diffusion information. The problem can be formalized as a graph construction problem from partially observed diffusion processes. It has been studied very recently in [38]. In our case, the originality comes either from the existence of different sources of observations or from the large impact of node contents in the network.

We will study how to combine graphs defined by networked data and graphs built from flat data to solve a given task. This is of major importance for information networks because, as said above, we will have to deal with multiple relations between entities (texts, spans of texts, ...) and also use textual data and vectorial data.

3.4. Prediction on Graphs and Scalability

As stated in the previous sections, graphs as complex objects provide a rich representation of data. Often enough the data is only partially available and the graph representation is very helpful in predicting the unobserved elements. We are interested in problems where the complete structure of the graph needs to be recovered and only a fraction of the links is observed. The link prediction problem falls into this category. We are also interested in the recommendation and link classification problems which can be seen as graphs where the structure is complete but some labels on the links (weights or signs) are missing. Finally we are also interested in labeling the nodes of the graph, with class or cluster memberships or with a real value, provided that we have (some information about) the labels for some of the nodes.

The semi-supervised framework will be also considered. A midterm research plan is to study how graph regularization models help for structured prediction problems. This question will be studied in the context of NLP tasks, as noted in Section 3.2, but we also plan to develop original machine learning algorithms that have a more general applicability. Inputs are networks whose nodes (texts) have to be labeled by structures. We assume that structures lie in some manifold and we want to study how labels can propagate in the network. One approach is to find a smooth labeling function corresponding to an harmonic function on both manifolds in input and output.

Scalability is one of the main issues in the design of new prediction algorithms working on networked data. It has gained more and more importance in recent years, because of the growing size of the most popular networked data that are now used by millions of people. In such contexts, learning algorithms whose computational complexity scales quadratically, or slower, in the number of considered data objects (usually nodes or edges, depending on the task) should be considered impractical.

These observations lead to the idea of using graph sparsification techniques in order to work on a part of the original network for getting results that can be easily extended and used for the whole original input. A sparsified version of the original graph can often be seen as a subset of the initial input, i.e. a suitably selected input subgraph which forms the training set (or, more in general, it is included in the training set). This holds even for the active setting. A simple example could be to find a spanning tree of the input graph, possibly using randomization techniques, with properties such that we are allowed to obtain interesting results for the initial graph dataset. We have started to explore this research direction for instance in [46].

At the level of mathematical foundations, the key issue to be addressed in the study of (large-scale) random networks also concerns the segmentation of network data into sets of independent and identically distributed observations. If we identify the data sample with the whole network, as it has been done in previous approaches [37], we typically end up with a set of observations (such as nodes or edges) which are highly interdependent and hence overly violate the classic i.i.d. assumption. In this case, the data scale can be so large and the range of correlations can be so wide, that the cost of taking into account the whole data and their dependencies is typically prohibitive. On the contrary, if we focus instead on a set of subgraphs independently drawn from a (virtually infinite) target network, we come up with a set of independent and identically distributed observations—namely the subgraphs themselves, where subgraph sampling is the underlying ergodic process [28]. Such an approach is one principled direction for giving novel statistical foundations to random network modeling. At the same time, because one shifts the focus from the whole network to a set of subgraphs, complexity issues can be restricted to the number of subgraphs and their size. The latter quantities can be controlled much more easily than the overall network size and dependence relationships, thus allowing to tackle scalability challenges through a radically redesigned approach.

Another way to tackle scalability problems is to exploit the inherent decentralized nature of very large graphs. Indeed, in many situations very large graphs are the abstract view of the digital activities of a very large set of users equipped with their own device. Nowadays, smartphones, tablets and even sensors have storage and computation power and gather a lot of data that serve to analytics, prediction, suggestion and personalized recommendation. Gathering all user data in large data centers is costly because it requires oversized infrastructures with huge energy consumption and large bandwith networks. Even though cloud architectures can optimize such infrastructures, data concentration is also prone to security leaks, lost of privacy and data governance for end users. The alternative we have started to develop in Magnet is to devise decentralized, private and personalized machine learning algorithms so that they can be deployed in the personal devices. The key challenges are therefore to learn in a collaborative way in a network of learners and to preserve privacy and control on personal data.

3.5. Beyond Homophilic Relationships

In many cases, algorithms for solving node classification problems are driven by the following assumption: linked entities tend to be assigned to the same class. This assumption, in the context of social networks, is known as homophily ([29], [39]) and involves ties of every type, including friendship, work, marriage, age, gender, and so on. In social networks, homophily naturally implies that a set of individuals can be parted into subpopulations that are more cohesive. In fact, the presence of homogeneous groups sharing common interests is a key reason for affinity among interconnected individuals, which suggests that, in spite of its simplicity, this principle turns out to be very powerful for node classification problems in general networks.

Recently, however, researchers have started to consider networked data where connections may also carry a negative meaning. For instance, disapproval or distrust in social networks, negative endorsements on the Web. Although the introduction of signs on graph edges appears like a small change from standard weighted graphs, the resulting mathematical model, called signed graphs, has an unexpectedly rich additional complexity. For example, their spectral properties, which essentially all sophisticated node classification algorithms rely on, are different and less known than those of graphs. Signed graphs naturally lead to a specific inference problem that we have discussed in previous sections: link classification. This is the problem of predicting signs of links in a given graph. In online social networks, this may be viewed as a form of sentiment analysis, since we would like to semantically categorize the relationships between individuals.

Another way to go beyond homophily between entities will be studied using our recent model of hypergraphs with bipartite hyperedges [41]. A bipartite hyperedge connects two ends which are disjoint subsets of nodes. Bipartite hyperedges is a way to relate two collections of (possibly heterogeneous) entities represented by nodes. In the NLP setting, while hyperedges can be used to model bags of words, bipartite hyperedges are associated with relationships between bags of words. But each end of bipartite hyperedges is also a way to represent complex entities, gathering several attribute values (nodes) into hyperedges viewed as records. Our hypergraph notion naturally extends directed and undirected weighted graph. We have defined a spectral theory for this new class of hypergraphs and opened a way to smooth labeling on sets of nodes. The weighting scheme allows to weigh the participation of each node to the relationship modeled by bipartite hyperedges accordingly to an equilibrium condition. This condition provides a competition between nodes in hyperedges and allows interesting modeling properties that go beyond homophily and similarity over nodes (the theoretical analysis of our hypergraphs exhibits tight relationships with signed graphs). Following this competition idea, bipartite hyperedges are like matches between two teams and examples of applications are team creation. The basic tasks we are interested in are hyperedge classification, hyperedge prediction, node weight prediction. Finally, hypergraphs also represent a way to summarize or compress large graphs in which there exists highly connected couples of (large) subsets of nodes.

4. Application Domains

4.1. Targeted Applications

Our main targeted applications are browsing, monitoring, recommending and mining in information networks. The learning tasks considered in the project such as node clustering, node and link classification and link prediction are likely to yield important improvements in these applications. Application domains cover social networks for cultural data and e-commerce, and biomedical informatics.

5. Highlights of the Year

5.1. Highlights of the Year

- We have been successful in many calls: ERC PoC project SOM (JAN RAMON leader), ANR project GRASP (PASCAL DENIS leader), ANR project PAMELA (MARC TOMMASI is the scientific coordinator), ANR project REM (PASCAL DENIS local leader), ADEME project MUST (JAN RAMON leader), Inria Associate Team LEGO (AURÉLIEN BELLET local leader).

- Scientific advances have been recognized by the community, in top ranked conferences and journals such as ICML, NIPS, JMLR, EMNLP, EACL and IJCAI.

5.1.1. Awards

- CHLOÉ BRAUD, who was supervised by PASCAL DENIS from 2012 to 2015, received the 2016 PhD Award from ATALA, the French NLP association.
- PAUL VANAESEBROUCK, who was supervised par AURÉLIEN BELLET and MARC TOMMASI, has reveived the "Grand Prix du stage de Recherche" from École Polytechnique Paris for his internship in MAGNET (see Section 7.1).

6. New Software and Platforms

6.1. CoRTex

Python library for noun phrase COreference Resolution in natural language TEXts
FUNCTIONAL DESCRIPTION

CoRTex is a LGPL-licensed Python library for Noun Phrase coreference resolution in natural language texts. This library contains implementations of various state-of-the-art coreference resolution algorithms, including those developed in our research. In addition, it provides a set of APIs and utilities for text pre-processing, reading the main annotation formats (ACE, CoNLL and MUC), and performing evaluation based on the main evaluation metrics (MUC, B-CUBED, and CEAF). As such, CoRTex provides benchmarks for researchers working on coreference resolution, but it is also of interest for developers who want to integrate a coreference resolution within a larger platform.

- Participants: Pascal Denis and David Chatel
- Contact: Pascal Denis
- URL: https://team.inria.fr/magnet/software/

6.2. Magneto

Python toolbox for generating and evaluating vector space representations for Natural Language Processing
FUNCTIONAL DESCRIPTION

Version 1.0 of Magneto contains preprocessing methods for texts in french and english. It includes classical methods for generating vector space representations: count based models, dimensionality reduction based methods and predictive methods (word2vec and Glove). For version 1.0, vector space representations can be evaluated on dedicated evaluation tasks such as similarity and analogy.

- Participants: Pascal Denis, Rémi Gilleron, Mikaela Keller, François Noyer and Nathalie Vauquier
- Contact: Pascal Denis
- URL: https://team.inria.fr/magnet/software/

7. New Results

7.1. Decentralized and Private Learning

In [13], we address the problem of decentralized minimization of pairwise functions of the data points, where these points are distributed over the nodes of a graph defining the communication topology of the network. This general problem finds applications in ranking, distance metric learning and graph inference, among others. We propose new gossip algorithms based on dual averaging which aims at solving such problems both in synchronous and asynchronous settings. The proposed framework is flexible enough to deal with constrained and regularized variants of the optimization problem. Our theoretical analysis reveals that the proposed algorithms preserve the convergence rate of centralized dual averaging up to an additive bias term. We present numerical simulations on Area Under the ROC Curve (AUC) maximization and metric learning problems which illustrate the practical interest of our approach.

In [19], we consider a set of learning agents in a collaborative peer-to-peer network, where each agent learns a *personalized model* according to its own learning objective. The question addressed in this paper is: how can agents improve upon their locally trained model by communicating with other agents that have similar objectives? We introduce and analyze two asynchronous gossip algorithms running in a fully decentralized manner. Our first approach, inspired from label propagation, aims to smooth pre-trained local models over the network while accounting for the confidence that each agent has in its initial model. In our second approach, agents jointly learn and propagate their model by making iterative updates based on both their local dataset and the behavior of their neighbors. Our algorithm for solving this challenging optimization problem relies on the Alternating Direction Method for Multipliers (ADMM).

In [20], we propose a decentralized protocol for a large set of users to privately compute averages over their joint data, which can later be used to learn more complex models. Our protocol can find a solution of arbitrary accuracy, does not rely on a trusted third party and preserves the privacy of users throughout the execution in both the honest-but-curious and malicious adversary models. Furthermore, we design a verification procedure which offers protection against malicious users joining the service with the goal of manipulating the outcome of the algorithm.

7.2. Natural Language Processing

In [12], we introduce a simple semi-supervised approach to improve implicit discourse relation identification. This approach harnesses large amounts of automatically extracted discourse connectives along with their arguments to construct new distributional word representations. Specifically, we represent words in the space of discourse connectives as a way to directly encode their rhetorical function. Experiments on the Penn Discourse Treebank demonstrate the effectiveness of these task-tailored representations in predicting implicit discourse relations. Our results indeed show that, despite their simplicity, these connective-based representations outperform various off-the-shelf word embeddings, and achieve state-of-the-art performance on this problem.

Along the PhD thesis of THIBAULT LIÉTARD, we are working on learning a similarity between text entities for the task of coreference resolution. Unlike indirect classification criteria often used in the literature, the similarity function naturally operates on pairs of mentions and several relevant objectives can be considered. For instance, we can learn the parameters of the similarity function such that the similarity of a given mention to its closest antecedent coreferent mention is larger than to any closer non-coreferential antecedent candidate. The resulting similarity scores can then be plugged into a greedy clustering procedure, or used to build a weighted graph of mentions to be clustered by spectral algorithms. For the representations of (pairs of) mentions on which the similarity function is learned, we consider both traditional linguistic features as well as external information about the general context of occurrence of the mentions using word embeddings.

Along the PhD thesis of MATHIEU DEHOUCK, we study the problem of cross-lingual dependency parsing, aiming at leveraging training data from different source languages to learn a parser in a target language. Specifically, this approach first constructs word vector representations that exploit structural (i.e., dependency-based) contexts but only considering the morpho-syntactic information associated with each word and its contexts. These delexicalized word embeddings, which can be trained on any set of languages and capture features shared accross languages are then used in combination with standard language-specific features to train a lexicalized parser in the target language. We evaluate our approach through experiments on a set of eight different languages that are part the Universal Dependencies Project. Our main results show that using such embeddings (monolingual or multilingual) achieves significant improvements over monolingual baselines. The work is submitted.

7.3. Edge Prediction in Networks

In [18] we address the problem of classifying the links of signed social networks given their full structural topology. In the problem of edge sign prediction, we are given a directed graph (representing a social network), and our task is to predict the binary labels of the edges (i.e., the positive or negative nature of the social relationships). Many successful heuristics for this problem are based on the troll-trust features, estimating at each node the fraction of outgoing and incoming positive/negative edges. We show that these heuristics can be understood, and rigorously analyzed, as approximators to the Bayes optimal classifier for a simple probabilistic model of the edge labels. We then show that the maximum likelihood estimator for this model approximately corresponds to the predictions of a label propagation algorithm run on a transformed version of the original social graph. Extensive experiments on a number of real-world datasets show that this algorithm is competitive against state-of-the-art classifiers in terms of both accuracy and scalability. Finally, we show that troll-trust features can also be used to derive online learning algorithms which have theoretical guarantees even when edges are adversarially labeled.

In [16], we address the problem of predicting connections between a set of data points. We focus on the *graph reconstruction* problem, where the prediction rule is obtained by minimizing the average error over all $n(n-1)/2$ possible pairs of the n nodes of a training graph. Our first contribution is to derive learning rates of order $O(\log n/n)$ for this problem, significantly improving upon the slow rates of order $O(1/\sqrt{n})$ established in the seminal work of [27]. Strikingly, these fast rates are universal, in contrast to similar results known for other statistical learning problems (e.g., classification, density level set estimation, ranking, clustering) which require strong assumptions on the distribution of the data. Motivated by applications to large graphs, our second contribution deals with the computational complexity of graph reconstruction. Specifically, we investigate to which extent the learning rates can be preserved when replacing the empirical reconstruction risk by a computationally cheaper Monte-Carlo version, obtained by sampling with replacement $B \ll n^2$ pairs of nodes. Finally, we illustrate our theoretical results by numerical experiments on synthetic and real graphs.

7.4. Mining Geotagged Social Data

Data generated on location-based social networks provide rich information on the whereabouts of urban dwellers. Specifically, such data reveal who spends time where, when, and on what type of activity (e.g., shopping at a mall, or dining at a restaurant). That information can, in turn, be used to describe city regions in terms of activity that takes place therein. For example, the data might reveal that citizens visit one region mainly for shopping in the morning, while another for dining in the evening. Furthermore, once such a description is available, one can ask more elaborate questions. For example, one might ask what features distinguish one region from another – some regions might be different in terms of the type of venues they host and others in terms of the visitors they attract. As another example, one might ask which regions are similar across cities. In [11], we present a method to answer such questions using publicly shared Foursquare data. Our analysis makes use of a probabilistic model, the features of which include the exact location of activity, the users who participate in the activity, as well as the time of the day and day of week the activity takes place. Compared to previous approaches to similar tasks, our probabilistic modeling approach allows us to make minimal assumptions about the data – which relieves us from having to set arbitrary parameters in our analysis (e.g., regarding the granularity of discovered regions or the importance of different features). We demonstrate how the model learned with our method can be used to identify the most likely and distinctive features of a geographical area, quantify the importance features used in the model, and discover similar regions across different cities. Finally, we perform an empirical comparison with previous work and discuss insights obtained through our findings. Our results were also presented through an interactive demo at the 25th World Wide Web Conference [21].

7.5. Learning from Non-iid Data

In [14] we deal with the generalization ability of classifiers trained from non-iid evolutionary-related data in which all training and testing examples correspond to leaves of a phylogenetic tree. For the realizable case, we prove PAC-type upper and lower bounds based on symmetries and matchings in such trees.

In [9], we studied learning problems where the performance criterion consists of an average over tuples (e.g., pairs or triplets) of observations rather than over individual observations, as in many learning problems involving networked data (e.g., link prediction), but also in metric learning and ranking. In this setting, the empirical risk to be optimized takes the form of a U-statistic, and its terms are highly dependent and thus violate the classic i.i.d. assumption. From a computational perspective, the calculation of such statistics is highly expensive even for a moderate sample size n, as it requires averaging $O(n^d)$ terms. We show that, strikingly, such empirical risks can be replaced by drastically computationally simpler Monte-Carlo estimates based on $O(n)$ terms only, usually referred to as incomplete U-statistics, without damaging the $O(1/\sqrt{n})$ learning rate of Empirical Risk Minimization (ERM) procedures. For this purpose, we establish uniform deviation results describing the error made when approximating a U-process by its incomplete version under appropriate complexity assumptions. Extensions to model selection, fast rate situations and various sampling techniques are also considered , as well as an application to stochastic gradient descent for ERM. Finally, numerical examples are displayed in order to provide strong empirical evidence that the approach we promote largely surpasses more naive subsampling techniques.

7.6. Adaptive Graph Construction

The efficiency of graph-based semi-supervised algorithms depends on the graph of instances on which they are applied. The instances are often in a vectorial form before a graph linking them is built. The construction of the graph relies on a metric over the vectorial space that help define the weight of the connection between entities. The classic choice for this metric is usually a distance measure or a similarity measure based on the euclidean norm. We claim that in some cases the euclidean norm on the initial vectorial space might not be the more appropriate to solve the task efficiently. In the work [17], we propose an algorithm that aims at learning the most appropriate vectorial representation for building a graph on which the task at hand is solved efficiently. In addition to experimental results showing the interest of such an approach, we define initial conditions under which the graph-based classification is ensured to perform optimally.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Cifre Clic and Walk (2013-2016)

Participants: Mikaela Keller [correspondent], Pauline Wauquier, Marc Tommasi.

We have a one to one cooperation with the Clic and Walk company that makes marketing surveys by consumers (called clicwalkers). The goal of the company is to understand the community of clicwalkers (40 thousands in one year) and its evolution with two objectives: the first one is to optimize the attribution of surveys to clicwalkers, and the second is to expand company's market to foreign countries. Social data can be obtained from social networks (G+, Facebook, ...) but there is no explicit network to describe the clicwalkers community. But users activity in answering surveys as well as server logs can provide traces of information diffusion, geolocalisation data, temporal data, sponsorship, etc. We study the problem of adaptive graph construction from the clicwalkers network. Node (users) classification and clustering algorithms are applied. For the problem of survey recommendations, the problem of teams constitution in a bipartite graph of users and surveys is studied. Random graph modeling and generative models of random graphs will be one step towards the prediction of the evolution of clicwalkers community.

8.1.2. ADEME

ADEME project MUST: Méthodologie d'exploitation des donnees d'usage des vehicules et d'identification de nouveaux Services pour les usagers et les territoires. Jan Ramon is the local PI at Inria of this project.

9. Partnerships and Cooperations

9.1. Regional Initiatives

Participation to the *Data Advanced data science and technologies* project (CPER Data). This project, leaded by DAVID SIMPLOT-RYL, is organized following three axes: internet of things, data science, high performance computing. MAGNET is involved in the data science axis to develop machine learning algorithms for big data, structured data and heterogeneous data.

9.2. National Initiatives

9.2.1. ANR Pamela (2016-2020)

Participants: MARC TOMMASI [correspondent], AURÉLIEN BELLET, RÉMI GILLERON, FABIO VITALE

The Pamela project aims at developing machine learning theories and algorithms in order to learn local and personalized models from data distributed over networked infrastructures. Our project seeks to provide first answers to modern information systems built by interconnecting many personal devices holding private user data in the search of personalized suggestions and recommendations. More precisely, we will focus on learning in a collaborative way with the help of neighbors in a network. We aim to lay the first blocks of a scientific foundation for these new types of systems, in effect moving from graphs of data to graphs of data and learned models. We argue that this shift is necessary in order to address the new constraints arising from the decentralization of information that is inherent to the emergence of big data. We will in particular focus on the question of learning under communication and privacy constraints. A significant asset of the project is the quality of its industrial partners, Snips and Mediego, who bring in their expertise in privacy protection and distributed computing as well as use cases and datasets. They will contribute to translate this fundamental research effort into concrete outcomes by developing personalized and privacy-aware assistants able to provide contextualized recommendations on small devices and smartphones. https://project.inria.fr/pamela/.

9.2.2. ANR JCJC GRASP (2016-2020)

Participants: PASCAL DENIS [correspondent], AURÉLIEN BELLET, RÉMI GILLERON, MIKAELA KELLER, MARC TOMMASI

The GRASP project aims at designing new graph-based Machine Learning algorithms that are better tailored to Natural Language Processing structured output problems. Focusing on semi-supervised learning scenarios, we will extend current graph-based learning approaches along two main directions: (i) the use of structured outputs during inference, and (ii) a graph construction mechanism that is more dependent on the task objective and more closely related to label inference. Combined, these two research strands will provide an important step towards delivering more adaptive (to new domains and languages), more accurate, and ultimately more useful language technologies. We will target semantic and pragmatic tasks such as coreference resolution, temporal chronology prediction, and discourse parsing for which proper Machine Learning solutions are still lacking. https://project.inria.fr/grasp/.

9.2.3. ANR-NFS REM (2016-2020)

With colleagues from the linguistics departments at Lille 3 and Neuchâtel (Switzerland), PASCAL DENIS is a member of another ANR project (REM), funded through the bilateral ANR-NFS Scheme. This project, co-headed by I. Depreatere (Lille 3) and M. Hilpert (Neufchâtel), proposes to reconsider the analysis of English modal constructions from a multidisciplinary perspective, combining insights from theoretical, psycho-linguistic, and computational approaches.

9.2.4. EFL (2010-2020)

PASCAL DENIS is an associate member of the Laboratoire d'Excellence *Empirical Foundations of Linguistics* (EFL), http://www.labex-efl.org/.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

ERC-PoC 713626 SOM "Statistical modeling for Optimization Mobility": This project aims at bringing to practice results from the project ERC-StG 240186 MiGraNT in the domain of mobility and mobile devices. In particular, a proof of concept will be made of graph mining approaches to learn predictive models and/or recommendation systems from collections of data distributed over a large number of devices (cars, smartphones, ...) while caring about privacy-friendliness.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

9.3.2.1. Sci-GENERATION (2013-2017)

Program: COST

Project acronym: Sci-GENERATION

Project title: Next Generation of Young Scientist: Towards a Contemporary Spirit of R&I.

Duration: 2013-2017

Coordinator: JAN RAMON is an MC member for belgium and a core group member

Other partners: More information on http://scigeneration.eu/en/participants.html

Abstract: Sci-Generation is a COST targeted network that addresses the challenges faced by next generation of researchers in Europe. We aim to improve the visibility, inclusion and success of excellent young researchers and research teams in European science and policy-making. We study and deliberate how changes in research funding opportunities and career perspectives can facilitate these improvements. We wish to promote new and emergent research topics, methods and management organisations. We are developing recommendations for EU science policy that will foster transformations at national and regional levels to promote scientific excellence and to establish a true European research area. (See http://scigeneration.eu).

9.3.2.2. TextLink (2014-2018)

Program: COST Action

Project acronym: TextLink

Project title: Structuring Discourse in Multilingual Europe

Duration: Apr. 2014 - Apr. 2018

Coordinator: Prof. Liesbeth Degand, Université Catholique de Louvain, Belgium. PASCAL DENIS is member of the Tools group.

Other partners: 26 EU countries and 3 international partner countries (Argentina, Brazil, Canada)

Abstract: Effective discourse in any language is characterized by clear relations between sentences and coherent structure. But languages vary in how relations and structure are signaled. While monolingual dictionaries and grammars can characterize the words and sentences of a language and bilingual dictionaries can do the same between languages, there is nothing similar for discourse. For discourse, however, discourse-annotated corpora are becoming available in individual languages. The Action will facilitate European multilingualism by (1) identifying and creating a portal into such resources within Europe - including annotation tools, search tools, and discourse-annotated corpora; (2) delineating the dimensions and properties of discourse annotation across corpora; (3) organizing these properties into a sharable taxonomy; (4) encouraging the use of this taxonomy in subsequent discourse annotation and in cross-lingual search and studies of devices that relate and structure discourse; and (5) promoting use of the portal, its resources and sharable taxonomy. With partners from across Europe, TextLink will unify numerous but scattered linguistic resources on discourse structure. With its resources searchable by form and/or meaning and a source of valuable correspondences, TextLink will enhance the experience and performance of human translators, lexicographers, language technology and language learners alike.

9.3.2.3. STAC (2011-2016)

Program: ERC Advanced Grant

Project acronym: STAC

Project title: Strategic conversation

Duration: Sep. 2011 - Aug. 2016

Coordinator: Nicholas Asher, CNRS, Université Paul Sabatier, IRIT (France)

Other partners: School of Informatics, Edinburgh University; Heriot Watt University, Edinburgh; Inria (PASCAL DENIS)

Abstract: STAC is a five year interdisciplinary project that aims to develop a new, formal and robust model of conversation, drawing from ideas in linguistics, philosophy, computer science and economics. The project brings a state of the art, linguistic theory of discourse interpretation together with a sophisticated view of agent interaction and strategic decision making, taking advantage of work on game theory.

9.4. International Initiatives

9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

9.4.1.1. RSS

Program: Inria North-European Labs

Project title: Rankings and Similarities in Signed graphs

Duration: late 2015 to late 2017

Partners: Aristides Gionis (Data Mining Group, Aalto University, Finland) and Mark Herbster (Centre for Computational Statistics and Machine Learning, University College London, UK)

Abstract: The project focuses on predictive analysis of networked data represented as signed graphs, where connections can carry either a positive or a negative semantic. The goal of this associate team is to devise novel formal methods and machine learning algorithms towards link classification and link ranking in signed graphs and assess their performance in both theoretical and practical terms.

9.4.1.2. LEGO

Title: LEarning GOod representations for natural language processing

International Partner (Institution - Laboratory - Researcher): University of California, Los Angeles (United States) - TEDS: Research group Theoretical and Empirical Data Science - Fei Sha

Start year: 2016

See also: https://team.inria.fr/lego/

Abstract: LEGO lies in the intersection of Machine Learning and Natural Language Processing (NLP). Its goal is to address the following challenges: what are the right representations for structured data and how to learn them automatically, and how to apply such representations to complex and structured prediction tasks in NLP? In recent years, continuous vectorial embeddings learned from massive unannotated corpora have been increasingly popular, but they remain far too limited to capture the complexity of text data as they are task-agnostic and fall short of modeling complex structures in languages. LEGO strongly relies on the complementary expertise of the two partners in areas such as representation/similarity learning, structured prediction, graph-based learning, and statistical NLP to offer a novel alternative to existing techniques. Specifically, we will investigate the following three research directions: (a) optimize the embeddings based on annotations so as to minimize structured prediction errors, (b) generate embeddings from rich language contexts represented as graphs, and (c) automatically adapt the context graph to the task/dataset of interest by learning a similarity between nodes to appropriately weigh the edges of the graph. By exploring these complementary research strands, we intend to push the state-of-the-art in several core NLP problems, such as dependency parsing, coreference resolution and discourse parsing.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

We invited Soravit Changpinyo (University of Southern California) in October, collaborating with Mathieu Dehouck, Pascal Denis and Aurélien Bellet on multi-task learning and transfer of word embeddings.

Jan Ramon collaborated with Wilhelmiina Hamalainen, who visited the magnet lab for 2 weeks. In particular, they worked on multiple hypothesis tests for regression and discretization problems.

Mark Herbster from University College London was invited for one week in January and collaborated with Fabio Vitale and Marc Tommasi on machine learning and similarity prediction in graphs.

Several international researchers have also been invited to give a talk at the Magnet seminar:

- Tim Vandercruys (Toulouse): “Modeling Meaning with Latent Factorization Models” (April)
- Soravit Changpinyo (University of Southern California): “Synthesized Classifiers for Zero-Shot Learning” (October)
- Thomas Kipf (University of Amsterdam): “Deep Learning on Graphs with Graph Convolutional Networks” (December)

9.5.1.1. Local Workshops

- Fabio Vitale organized the workshop Graph-based Learning and Graph Mining.
- Pascal Denis organized the Workshop on Argumentation Mining.

9.5.2. Visits to International Teams

In March, April and May Fabio Vitale visited the Department of Computer Science of the University of Milan, collaborating with Prof. Nicolò Cesa-Bianchi and Prof. Claudio Gentile.

In July, Aurélien Bellet and Pascal Denis visited the Department of Computer Science of the University of California (Los Angeles), collaborating with Prof Fei Sha.

In September, Mathieu Dehouck visited the Department of Computer Science of the University of California (Los Angeles), collaborating with Prof Fei Sha.

Since September, Fabio Vitale is working at the department of computer science of Aalto University, Helsinki (Finland), in the DMG group (http://research.ics.aalto.fi/dmg/index.shtml) led by Prof. Aristides Gionis.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

Aurélien Bellet co-organized the workshop Private Multi-Party Machine Learning @ NIPS 2016. [0]

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Pascal Denis served as co-chair of the Polaris Colloquium, a monthly Guest Lecture series in Computer Science and Signal Processing, co-sponsored by Inria Lille–Nord Europe and University of Lille CRIStAL Lab.

[0] https://pmpml.github.io/PMPML16/

10.1.2.2. Member of the Conference Program Committees

AURÉLIEN BELLET served as PC member for IJCAI 2016, ICML 2016, NIPS 2016 and AISTATS 2017.

PASCAL DENIS served as Senior PC member for IJCAI 2016. He was PC member for ACL 2016, AAAI 2016, CAp 2016, EMNLP 2016, and NAACL 2016.

RÉMI GILLERON served as PC member for NIPS 2016 and AISTATS 2017.

JAN RAMON served as PC member for AISTATS 2016 and 2017, BAI workshop @ IJCAI 2017, ieee big data 2016, BNAIC 2016, DS 2016, ECAI 2016, ECML-PKDD 2016, IEEE ICDM 2016, ICHI 2016, IJCAI 2016, ILP 2016, ISMIS 2017, KDD 2016, MLG 2016, NIPS 2016, PMPML workshop @ NIPS 2016, SSDM workshop @ ECML-PKDD 2016.

MARC TOMMASI served as PC member for NIPS 2016, ICML 2016, IJCAI 2016, EGC 2017.

FABIEN TORRE served as PC member for EGC 2017, workshop CluCo 2017, AAFD & SFC 2016, CNIA 2016

FABIO VITALE served as PC member for AISTATS 2017 and ECMLPKDD 2016, NIPS 2016.

10.1.2.3. Reviewer

GÉRAUD LE FALHER was reviewer for ECMLPKDD 2016, NIPS 2016 and IJCAI 2016. MIKAELA KELLER was a reviewer for ICLR 2016 and ICML 2016.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

JAN RAMON is member of the editorial boards of data mining and knowledge discovery, machine learning and guest editorial board of ECML-PKDD 2016.

10.1.3.2. Reviewer - Reviewing Activities

AURÉLIEN BELLET was reviewer for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) and IEEE Transactions on Cybernetics (TCYB).

JAN RAMON: data mining and knowledge discovery (dami) 5; knowledge and information systems (kais): 4; machine learning :3; neurocomputing: 2; artificial intelligence: 2; plos one : 1

FABIO VITALE was reviewer for the journal EJOR (European Journal of Operational Research - Elsevier), and the journal Internet Mathematics.

GÉRAUD LE FALHER was reviewer for the International Journal of E-Planning Research.

10.1.4. Invited Talks

AURÉLIEN BELLET was invited the Learning, Privacy, and Mobile Data Workshop at Google Research Seattle, [0] the Statistical Machine Learning (SMILE) seminar in Paris, [0] the Workshop on Distributed Machine Learning (Télécom ParisTech), [0] and at the company EURA NOVA. [0]

PASCAL DENIS was invited to the STL seminar, Université Lille 3 (February 2016).

10.1.5. Scientific Expertise

PASCAL DENIS was reviewer for Flanders Research Foundation (FWO, Belgium) and the Brussels Institute for Research and Innovation (Innoviris).

JAN RAMON was reviewer of H2020 projects.

10.1.6. Research Administration

FABIEN TORRE is in the board of the national evaluation committee for teaching and research in computer science (CNU 27)

[0] https://sites.google.com/site/learningprivacymobiledata/
[0] https://sites.google.com/site/smileinparis/
[0] http://machinelearningforbigdata.telecom-paristech.fr/fr/article/workshop-friday-novembre-25-distributed-machine-learning
[0] http://euranova.eu

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence MIASHS: FABIO VITALE, Introduction à l'algorithmique, 66h, L2, Université Lille 3.

Licence MIASHS: MARC TOMMASI, Réseaux, 28h, L1, Université Lille 3.

Licence MIASHS: RÉMI GILLERON, Traitement de données, 24h, L1, Université Lille 3.

Licence MIASHS: GÉRAUD LE FALHER, Traitement de données, 36h, L1, Université Lille 3.

Licence MIASHS: MATHIEU DEHOUCK et THIBAULT LIÉTARD, Projet informatique de traitement de données en SHS, 20h, L2, Université Lille 3.

Licence MIASHS: MIKAELA KELLER, Codage et représentation de l'information, 24h, L1, Université Lille 3.

Licence SoQ (SHS): FABIEN TORRE, Traitement de contenus textuels, 24h, L3, Université Lille 3.

Licence SoQ (SHS): RÉMI GILLERON, Algorithmique de graphes, 24h, L3, Université Lille 3.

Licence SHS: MIKAELA KELLER, Langages du Web, 24h, L3, Université Lille 3.

Licence SHS: MIKAELA KELLER, Représentation numérique de l'information, 24h, L3, Université Lille 3.

Licence économie gestion: RÉMI GILLERON, Traitement de données et documents, 24h, L1, Université Lille 3.

Licence MARC TOMMASI, MIKAELA KELLER C2i, université Lille 3.

Master MOCAD: PASCAL DENIS co-taught the Machine Learning and Decision under Uncertainty class, 37,5h, Université Lille 1.

Master MIASHS: RÉMI GILLERON et FABIEN TORRE, Web et référencement, 24h, M1, Université Lille 3.

Master MIASHS: GÉRAUD LE FALHER, Web et réseaux, 24h, M1, Université Lille 3.

Master MIASHS: MIKAELA KELLER, Programmation et bases de données, 24h, M1, Université Lille 3.

Master LTTAC: FABIEN TORRE, Algorithmique des textes – Javascript, 36h, M1, Université Lille 3.

Master ID: FABIEN TORRE, information structurée, 20h, M2, Université Lille 3.

Master ID: FABIEN TORRE, programmation Web, 20h, M2, Université Lille 3.

Master / Master Spécialisé Big Data: AURÉLIEN BELLET, Advanced Machine Learning, 25.5h, Télécom ParisTech.

Formation continue (Certificat d'Études Spécialisées Data Scientist): AURÉLIEN BELLET, Supervised Learning and Support Vector Machines, 10h, Télécom ParisTech.

Formation continue: AURÉLIEN BELLET, Graph Mining, 3h, Télécom ParisTech pour Allianz.

E-learning

SPOC: MARC TOMMASI, RÉMI GILLERON and ALAIN PREUX: Culture numérique, 5 semesters at the bachelor level, Moodle, Lille 3 university, more than 7000 students.

Pedagogical resources: texts, videos, quizz and exercices available on http://culturenumerique.univ-lille3.fr/, creative commons.

10.2.2. Supervision

PhD in progress: GÉRAUD LE FALHER, Machine Learning in Signed Graphs, Inria Lille – Nord Europe, since Oct. 2014, MARC TOMMASI, FABIO VITALE and CLAUDIO GENTILE (University of Insubria, Italy).

Phd in progress: David Chatel, Semi-supervised spectral clustering since Sep 2012, Marc Tommasi and Pascal Denis.

Phd in progress: Mathieu Dehouck, Graph-based Learning for Multi-lingual and Multi-domain Dependency Parsing, since Oct 2015, Pascal Denis and Marc Tommasi.

Phd in progress: Pauline Wauquier, Recommendation in Information Networks, since Dec 2013, Mikaela Keller and Marc Tommasi.

Phd in progress: Thibault Liétard, Adaptive Graph Learning with Applications to Natural Language Processing, Aurélien Bellet, Pascal Denis and Rémi Gilleron.

Master: Thibault Liétard, Metric learning for Graph-based coreference resolution, ENS Rennes, co-supervised by Aurélien Bellet and Pascal Denis.

Master: Paul Vanaesebrouck, Decentralized Machine Learning on Graphs, Ecole Polytechnique, co-supervised by Aurélien Bellet and Marc Tommasi.

Master: Pierre Dellenbach, Private learning of heavily distributed data, Ecole Polytechnique, co-supervised by Aurélien Bellet and Jan Ramon.

Master: Robin Vogel, Learning to Rank Rare Instances, ENSAE, co-supervised by Aurélien Bellet, Stéphan Clémençon et Anne Sabourin (Télécom ParisTech), et Stéphane Gentric (Morpho).

10.2.3. Juries

- Aurélien Bellet was member of the recruitment committee for MdC in Computer Science at Télécom Saint-Etienne.
- Pascal Denis et Marc Tommasi were members of the recruitment committee for MdC in Computer Science at Université Lille 3.
- Pascal Denis was member of the Commission Emploi Rechercher (CER) at Inria Lille – Nord Europe.
- Rémi Gilleron was member of the PhD committees of Hugo Louche (Examinateur) and Tom Sebastian (Président).
- Marc Tommasi was member of the habilitation committee of Albert Biffet (Rapporteur).
- Marc Tommasi was member of the PhD committees of Guillaume Rabusseau (Rapporteur), Raphaël Puget (Rapporteur), Michaël Perrot (Examinateur), Hadrien Glaude (Examinateur).
- Marc Tommasi was head of the jury for the recruitment committee of Junior Research Scientists (CR1/CR2) at Inria Lille.

10.3. Popularization

Marc Tommasi presented the Magnet team at the EuraTechnologies ICT innovation ecosystem (https://www.inria.fr/centre/lille/agenda/r-dv-du-plateau-inria-apprentissage-automatique-et-reseaux-d-information).

Aurélien Bellet presented some of his work at the popularization seminar “30 minutes de sciences” and at an Assemblée Générale of Inria Lille - Nord Europe.

Thibault Liétard has participated to the “chercheur itinérant” initiative (https://www.inria.fr/centre/lille/recherche/sciences-pour-tous2/mediation/chercheurs-itinerants-au-lycee-2016).

Rémi Gilleron has participated at the forum “F O O R <: Forum Ouvert Oeuvres et Recherches” where he presented the research work done with Pascal Denis for the artwork “This is Major Tom to Ground Control”

11. Bibliography

Major publications by the team in recent years

[1] A. FRENO, M. KELLER, M. TOMMASI.*Fiedler Random Fields: A Large-Scale Spectral Approach to Statistical Network Modeling*, in "Neural Information Processing Systems (NIPS)", Lake Tahoe, United States, Advances in Neural Information Processing Systems, MIT Press, December 2012, vol. 25, https://hal.inria.fr/hal-00750345.

[2] O. KUŽELKA, Y. WANG, J. RAMON.*Bounds for Learning from Evolutionary-Related Data in the Realizable Case*, in "International Joint Conference on Artificial Intelligence (IJCAI)", New York, United States, Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) 2016, July 2016, https://hal.archives-ouvertes.fr/hal-01422033.

[3] E. LASSALLE, P. DENIS.*Improving pairwise coreference models through feature space hierarchy learning*, in "ACL 2013 - Annual meeting of the Association for Computational Linguistics", Sofia, Bulgaria, Association for Computational Linguistics, August 2013, https://hal.inria.fr/hal-00838192.

[4] E. LASSALLE, P. DENIS.*Joint Anaphoricity Detection and Coreference Resolution with Constrained Latent Structures*, in "AAAI Conference on Artificial Intelligence (AAAI 2015)", Austin, Texas, United States, Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI 2015), January 2015, https://hal.inria.fr/hal-01205189.

[5] G. PAPA, S. CLÉMENÇON, A. BELLET.*On Graph Reconstruction via Empirical Risk Minimization: Fast Learning Rates and Scalability*, in "Annual Conference on Neural Information Processing Systems (NIPS 2016)", Barcelone, Spain, December 2016, https://hal.inria.fr/hal-01367546.

[6] C. PELEKIS, J. RAMON, Y. WANG.*Hölder-type inequalities and their applications to concentration and correlation bounds*, in "Indagationes Mathematicae", 2016.

[7] T. RICATTE, R. GILLERON, M. TOMMASI.*Hypernode Graphs for Spectral Learning on Binary Relations over Sets*, in "ECML/PKDD - 7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Nancy, France, Machine Learning and Knowledge Discovery in Databases, September 2014, Paper accepted for publication at ECML/PKDD 2014, https://hal.inria.fr/hal-01017025.

Publications of the year

Articles in International Peer-Reviewed Journal

[8] A. BELLET, J. F. BERNABEU, A. HABRARD, M. SEBBAN.*Learning Discriminative Tree Edit Similarities for Linear Classification - Application to Melody Recognition*, in "Neurocomputing", 2016, vol. 214, p. 155-161 [*DOI :* 10.1016/J.NEUCOM.2016.06.006], https://hal.archives-ouvertes.fr/hal-01330492.

[9] S. CLÉMENÇON, I. COLIN, A. BELLET.*Scaling-up Empirical Risk Minimization: Optimization of Incomplete U-statistics*, in "Journal of Machine Learning Research (JMLR)", 2016, vol. 17, n° 76, p. 1-36, https://hal.inria.fr/hal-01327662.

[10] E. MAES, P. KELCHTERMANS, W. BITTREMIEUX, K. DE GRAVE, S. DEGROEVE, J. HOOYBERGHS, I. MERTENS, G. BAGGERMAN, J. RAMON, K. LAUKENS, L. MARTENS, D. VALKENBORG.*Designing biomedical proteomics experiments: state-of-the-art and future perspectives*, in "Expert Review of Proteomics", 2016 [*DOI :* 10.1586/14789450.2016.1172967], https://hal.inria.fr/hal-01431414.

[11] E. ÇELIKTEN, G. C. LE FALHER, M. C. MATHIOUDAKIS.*Modeling Urban Behavior by Mining Geotagged Social Data*, in "IEEE Transactions on Big Data", December 2016, 14 [*DOI :* 10.1109/TBDATA.2016.2628398], https://hal.inria.fr/hal-01406676.

International Conferences with Proceedings

[12] C. BRAUD, P. DENIS.*Learning Connective-based Word Representations for Implicit Discourse Relation Identification*, in "Empirical Methods on Natural Language Processing", Austin, United States, November 2016, https://hal.inria.fr/hal-01397318.

[13] I. COLIN, A. BELLET, J. SALMON, S. CLÉMENÇON.*Gossip Dual Averaging for Decentralized Optimization of Pairwise Functions*, in "International Conference on Machine Learning (ICML 2016)", New York, United States, June 2016, https://hal.inria.fr/hal-01329315.

[14] O. KUŽELKA, Y. WANG, J. RAMON.*Bounds for Learning from Evolutionary-Related Data in the Realizable Case*, in "International Joint Conference on Artificial Intelligence (IJCAI)", New York, United States, Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI) 2016, July 2016, https://hal.archives-ouvertes.fr/hal-01422033.

[15] Z. LU, D. GUO, A. B. GARAKANI, K. LIU, A. MAY, A. BELLET, L. FAN, M. COLLINS, B. KINGSBURY, M. PICHENY, F. SHA.*A Comparison Between Deep Neural Nets and Kernel Acoustic Models for Speech Recognition*, in "IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2016)", Shanghai, China, March 2016, https://hal.inria.fr/hal-01329772.

[16] G. PAPA, S. CLÉMENÇON, A. BELLET.*On Graph Reconstruction via Empirical Risk Minimization: Fast Learning Rates and Scalability*, in "Annual Conference on Neural Information Processing Systems (NIPS 2016)", Barcelone, Spain, December 2016, https://hal.inria.fr/hal-01367546.

Conferences without Proceedings

[17] P. WAUQUIER, M. KELLER.*A Metric Learning Approach for Graph-Based Label Propagation*, in "Workshop track of ICLR 2016", San Juan, Puerto Rico, May 2016, https://hal.inria.fr/hal-01427287.

Research Reports

[18] G. LE FALHER, N. CESA-BIANCHI, C. GENTILE, F. VITALE.*On the Troll-Trust Model for Edge Sign Prediction in Social Networks*, Inria Lille, 2016, https://hal.inria.fr/hal-01425137.

[19] P. VANHAESEBROUCK, A. BELLET, M. TOMMASI.*Decentralized Collaborative Learning of Personalized Models over Networks*, Inria Lille, October 2016, https://hal.inria.fr/hal-01383544.

Other Publications

[20] P. DELLENBACH, J. RAMON, A. BELLET.*A Decentralized and Robust Protocol for Private Averaging over Highly Distributed Data*, December 2016, NIPS 2016 workshop on Private Multi-Party Machine Learning, Poster, https://hal.inria.fr/hal-01384148.

[21] E. ÇELIKTEN, G. LE FALHER, M. MATHIOUDAKIS.*"What Is the City but the People?" Exploring Urban Activity Using Social Web Traces*, 25th World Wide Web Conference, Demo Track, April 2016, 25th World Wide Web Conference, Demo Track, Poster [*DOI :* 10.1145/2872518.2901922], https://hal.inria.fr/hal-01295344.

References in notes

[22] A. ALEXANDRESCU, K. KIRCHHOFF.*Graph-based learning for phonetic classification*, in "IEEE Workshop on Automatic Speech Recognition & Understanding, ASRU 2007, Kyoto, Japan, December 9-13, 2007", 2007, p. 359-364.

[23] M.-F. BALCAN, A. BLUM, P. P. CHOI, J. LAFFERTY, B. PANTANO, M. R. RWEBANGIRA, X. ZHU.*Person Identification in Webcam Images: An Application of Semi-Supervised Learning*, in "ICML2005 Workshop on Learning with Partially Classified Training Data", 2005.

[24] M. BELKIN, P. NIYOGI.*Towards a Theoretical Foundation for Laplacian-Based Manifold Methods*, in "Journal of Computer and System Sciences", 2008, vol. 74, n^o 8, p. 1289-1308.

[25] A. BELLET, A. HABRARD, M. SEBBAN.*A Survey on Metric Learning for Feature Vectors and Structured Data*, in "CoRR", 2013, vol. abs/1306.6709.

[26] A. BELLET, A. HABRARD, M. SEBBAN.*Metric Learning*, Morgan & Claypool Publishers, 2015.

[27] G. BIAU, K. BLEAKLEY.*Statistical Inference on Graphs*, in "Statistics & Decisions", 2006, vol. 24, p. 209–232.

[28] P. J. BICKEL, A. CHEN.*A nonparametric view of network models and Newman–Girvan and other modularities*, in "Proceedings of the National Academy of Sciences", 2009, vol. 106, p. 21068–21073.

[29] P. BLAU.*Inequality and Heterogeneity: A Primitive Theory of Social Structure*, MACMILLAN Company, 1977, http://books.google.fr/books?id=jvq2AAAAIAAJ.

[30] C. BRAUD, P. DENIS.*Combining Natural and Artificial Examples to Improve Implicit Discourse Relation Identification*, in "coling", Dublin, Ireland, August 2014, https://hal.inria.fr/hal-01017151.

[31] H. CHANG, D.-Y. YEUNG.*Graph Laplacian Kernels for Object Classification from a Single Example*, in "Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2", Washington, DC, USA, CVPR '06, IEEE Computer Society, 2006, p. 2011–2016, http://dx.doi.org/10.1109/CVPR.2006.128.

[32] D. CHATEL, P. DENIS, M. TOMMASI.*Fast Gaussian Pairwise Constrained Spectral Clustering*, in "ECML/PKDD - 7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Nancy, France, September 2014, p. 242 - 257 [*DOI :* 10.1007/978-3-662-44848-9_16], https://hal.inria.fr/hal-01017269.

[33] D. DAS, S. PETROV.*Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections*, in "ACL", 2011, p. 600-609.

[34] P. DENIS, P. MULLER.*Predicting globally-coherent temporal structures from texts via endpoint inference and graph decomposition*, in "IJCAI-11 - International Joint Conference on Artificial Intelligence", Barcelone, Espagne, 2011, http://hal.inria.fr/inria-00614765.

[35] E. R. FERNANDES, U. BREFELD.*Learning from Partially Annotated Sequences*, in "ECML/PKDD", 2011, p. 407-422.

[36] A. B. GOLDBERG, X. ZHU.*Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization*, in "Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing", Stroudsburg, PA, USA, TextGraphs-1, Association for Computational Linguistics, 2006, p. 45–52, http://dl.acm.org/citation.cfm?id=1654758.1654769.

[37] A. GOLDENBERG, A. X. ZHENG, S. E. FIENBERG.*A Survey of Statistical Network Models*, Foundations and trends in machine learning, Now Publishers, 2010, http://books.google.fr/books?id=gPGgcOf95moC.

[38] M. GOMEZ-RODRIGUEZ, J. LESKOVEC, A. KRAUSE.*Inferring networks of diffusion and influence*, in "Proc. of KDD", 2010, p. 1019-1028.

[39] M. MCPHERSON, L. S. LOVIN, J. M. COOK.*Birds of a Feather: Homophily in Social Networks*, in "Annual Review of Sociology", 2001, vol. 27, n^{o} 1, p. 415–444, http://dx.doi.org/10.1146/annurev.soc.27.1.415.

[40] A. NENKOVA, K. MCKEOWN.*A Survey of Text Summarization Techniques*, in "Mining Text Data", Springer, 2012, p. 43-76.

[41] T. RICATTE, R. GILLERON, M. TOMMASI.*Hypernode Graphs for Spectral Learning on Binary Relations over Sets*, in "ECML/PKDD - 7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Nancy, France, Machine Learning and Knowledge Discovery in Databases, September 2014, https://hal.inria.fr/hal-01017025.

[42] H. SHIN, K. TSUDA, B. SCHÖLKOPF.*Protein functional class prediction with a combined graph*, in "Expert Syst. Appl.", March 2009, vol. 36, n^{o} 2, p. 3284–3292, http://dx.doi.org/10.1016/j.eswa.2008.01.006.

[43] S. SINGH, A. SUBRAMANYA, F. C. N. PEREIRA, A. MCCALLUM.*Large-Scale Cross-Document Coreference Using Distributed Inference and Hierarchical Models*, in "ACL", 2011, p. 793-803.

[44] M. SPERIOSU, N. SUDAN, S. UPADHYAY, J. BALDRIDGE.*Twitter Polarity Classification with Label Propagation over Lexical Links and the Follower Graph*, in "Proceedings of the First Workshop on Unsupervised Methods in NLP", Edinburgh, Scotland, 2011.

[45] A. SUBRAMANYA, S. PETROV, F. C. N. PEREIRA.*Efficient Graph-Based Semi-Supervised Learning of Structured Tagging Models*, in "EMNLP", 2010, p. 167-176.

[46] F. VITALE, N. CESA-BIANCHI, C. GENTILE, G. ZAPPELLA.*See the Tree Through the Lines: The Shazoo Algorithm*, in "Proc of NIPS", 2011, p. 1584-1592.

[47] L. WANG, S. N. KIM, T. BALDWIN.*The Utility of Discourse Structure in Identifying Resolved Threads in Technical User Forums*, in "COLING", 2012, p. 2739-2756.

[48] K. K. YUZONG LIU.*Graph-Based Semi-Supervised Learning for Phone and Segment Classification*, in "Proceedings of Interspeech", Lyon, France, 2013.

[49] X. ZHU, Z. GHAHRAMANI, J. LAFFERTY.*Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, in "Proc. of ICML", 2003, p. 912-919.

Project-Team MEPHYSTO

Quantitative methods for stochastic models in physics

IN PARTNERSHIP WITH:
Université Libre de Bruxelles

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Numerical schemes and simulations

Table of contents

Project-Team MEPHYSTO

Creation of the Team: 2014 January 01, updated into Project-Team: 2016 July 01

Keywords:

Computer Science and Digital Science:

6. - Modeling, simulation and control
6.1. - Mathematical Modeling
6.1.1. - Continuous Modeling (PDE, ODE)
6.1.2. - Stochastic Modeling (SPDE, SDE)
6.1.4. - Multiscale modeling
6.2. - Scientific Computing, Numerical Analysis & Optimization
6.2.1. - Numerical analysis of PDE and ODE
6.2.2. - Numerical probability
6.2.3. - Probabilistic methods

Other Research Topics and Application Domains:

3.3.1. - Earth and subsoil
5.5. - Materials
9.4.2. - Mathematics
9.4.3. - Physics

1. Members

Research Scientists

Antoine Gloria [Team leader, ULB & Inria, Professor & Researcher, HDR]
Guillaume Dujardin [Inria, Researcher]
Marielle Simon [Inria, Researcher]

Faculty Members

Denis Bonheure [ULB, Professor]
Stephan de Bievre [Univ. Lille I, Professor, HDR]

PhD Students

Mitia Duerinckx [ULB]
Pierre Mennuni [Univ. Lille I, from Oct 2016]

Post-Doctoral Fellows

Antoine Benoit [ULB]
Jean Baptiste Casteras [ULB]
Christopher Shirley [ULB]

Visiting Scientist

Milton Jara Valenzuela [IMPA, from Oct 2016 until Nov 2016]

Administrative Assistant

Karine Lewandowski [Inria]

2. Overall Objectives

2.1. Presentation and overall objectives

In the context of the construction of the European landscape of research, Inria and ULB (Université Libre de Bruxelles) signed in 2013 an agreement to foster joint research teams on topics of mutual interests. The team MEPHYSTO, a joint project of Inria, the Université Lille 1 and CNRS, and the Université Libre de Bruxelles, is the first such collaboration, in applied mathematics. It operates in two locations: Lille and Brussels.

The main objective of the team is to develop mathematical and numerical tools to study in a quantitative way some specific physical models which display random and/or multiscale features. The emphasis is put on the interplay between analysis, probability, and numerics.

We focus our efforts on two prototypical examples: stochastic homogenization and the Schrödinger equations.

2.2. Scientific context

Whereas many models in physics involve randomness, they behave deterministically in suitable asymptotic regimes when stochastic effects average out. The qualitative and quantitative understanding of this deterministic behavior is the main challenge of this project.

From a mathematical point of view, our main fields of interest are stochastic homogenization of PDEs and random or deterministic one-dimensional nonlinear Schrödinger equations. These topics involve two challenges identified in the strategic plan of Inria “Objectif 2020": randomness and multiscale modeling.

From a physical point of view, the problems we shall consider find their origin in

- the statistical physics of random polymer-chain networks;
- light propagation in optical fibers.

Stochastic homogenization

Homogenization is a theory which deals with oscillations in PDEs. Let D be a smooth bounded domain of $\mathbb{R}^d$. The starting point is the fact that for linear elliptic equations, the oscillations of the weak solution $u_\varepsilon \in H^1_0(D)$ of

$$-\nabla \cdot A_\varepsilon \nabla u_\varepsilon \;=\; f \tag{3}$$

for some suitable r. h. s. f are a (nonlinear) function of the oscillations of A_ε. In particular, if A_ε oscillates at scale $\varepsilon > 0$, one expects u_ε to display oscillations at scale ε, and to be close to some function which does not oscillate if in addition $\varepsilon \ll 1$. This is the case when A_ε is the ε-rescaled version of a periodic function A. Then A_ε is ε-periodic, and there exists some fixed matrix A_{hom} depending only on A (and not on f), such that u_ε behaves as $u_{\mathrm{hom}} \in H^1_0(D)$, the weak solution of

$$-\nabla \cdot A_{\mathrm{hom}} \nabla u_{\mathrm{hom}} \;=\; f. \tag{4}$$

The homogenized coefficients A_{hom} are characterized by the so-called correctors ϕ_ξ in direction $\xi \in \mathbb{R}^d$, distributional solutions in $\mathbb{R}^d$ of

$$-\nabla \cdot A(\xi + \nabla \phi_\xi) \;=\; 0. \tag{5}$$

In the periodic case, these correctors are well-behaved by standard PDE theory. The convergence of u_ε to u_{hom} is illustrated on Figure 1 (periodic checkerboard on the left, random checkerboard on the right), where the isolines of the solutions to (1) and (2) (with $f \equiv 1$ on the unit square) are plotted for several values of ε — the convergence of u_ε to u_{hom} is weak in $H^1(D)$. Yet, naturally-occuring structures are rarely periodic. If instead of considering some periodic A, we consider some random A, the story is different, cf. Figure 1 for results on the random checkerboard. In the early period of stochastic homogenization, in the seventies, it was not clear if just the ergodicity and stationarity of the coefficients and ellipticity were enough to prove convergence of u_ε almost surely and identify the limit u_{hom}. The meaning to give to (3) was indeed quite unclear (the equation is posed on the whole space). It was a surprise, therefore, that this was possible with random coefficients, and that stochastic homogenization was indeed a new type of *qualitative* ergodic theory ([75], [71]). The following natural question, asked more than thirty years ago, is whether one can develop an associated *quantitative* ergodic theory.

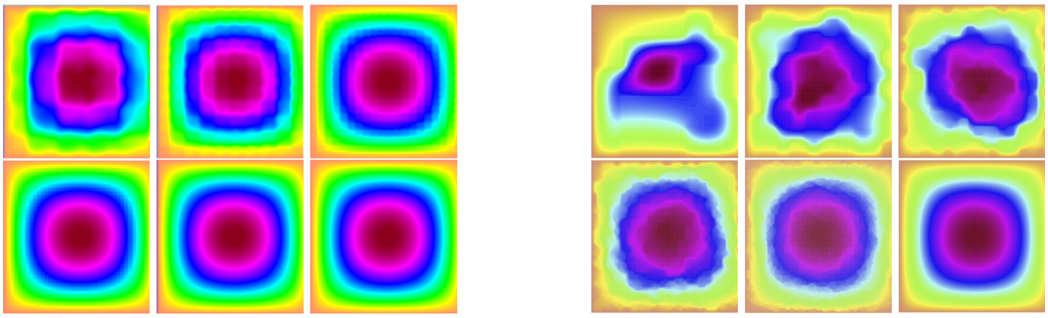


Figure 1. Solution u_ε for $\varepsilon = 1/5, 1/10, 1/20, 1/40, 1/80$ and solution u_{hom}, periodic case (left) and random case (right)

One of our initial motivations to develop a quantitative stochastic homogenization theory is the derivation of nonlinear elasticity from polymer physics, which is presented in the research program and application section. We plan to develop a complete quantitative theory of stochastic homogenization of elliptic equations. In particular we aim at quantifying how well u_{hom} approximates u_ε, and at identifying the asymptotic law of the solution u_ε in function of the law of A.

Schrödinger equations

The linear Schrödinger equation, with an appropriate choice of geometry and boundary conditions, has been central to the description of all non-relativistic quantum mechanical systems for almost a century now. In addition, its nonlinear variant arises in the mean field description of Bose-Einstein condensates, where it is known as the Gross-Pitaevskii equation, but also in nonlinear classical optics, and in particular in fiber optics. The quantitative and qualitative description of its solutions (for both the evolution and the stationary equations), their time-asymptotic behavior, their stability or instability in terms of the parameters of the initial conditions and/or the potentials and boundary conditions continue to pose numerous physical and mathematical problems (see [76] and [43] for general references).

In view of our collaboration with the Lille laser physics laboratory PhLAM, we will focus more particularly on the one-dimensional nonlinear Schrödinger equation (NLS). Indeed, (NLS) drives the envelope of the propagation of a laser pulse in a Kerr medium, such as an optical fiber [76]. Many phenomena on (NLS) (and variants thereof, with higher order derivatives, various types of initial conditions, external fields, etc.) are put in evidence by physical experiments at PhLAM, are not fully understood, and raise exciting questions from the numerical and analytical perspectives.

The same type of equation also describes Bose-Einstein condensates, for which questions related to Anderson localization are also of interest theoretically and experimentally at PhLAM.

3. Research Program

3.1. From statistical physics to continuum mechanics

Whereas numerical methods in nonlinear elasticity are well-developed and reliable, constitutive laws used for rubber in practice are phenomenological and generally not very precise. On the contrary, at the scale of the polymer-chain network, the physics of rubber is very precisely described by statistical physics. The main challenge in this field is to understand how to derive macroscopic constitutive laws for rubber-like materials from statistical physics.

At the continuum level, rubber is modelled by an energy E defined as the integral over a domain D of $\mathbb{R}^d$ of some energy density W depending only locally on the gradient of the deformation u: $E(u) = \int_D W(\nabla u(x))dx$. At the microscopic level (say 100nm), rubber is a network of cross-linked and entangled polymer chains (each chain is made of a sequence of monomers). At this scale the physics of polymer chains is well-understood in terms of statistical mechanics: monomers thermally fluctuate according to the Boltzmann distribution [63]. The associated Hamiltonian of a network is typically given by a contribution of the polymer chains (using self-avoiding random bridges) and a contribution due to steric effects (rubber is packed and monomers are surrounded by an excluded volume). The main challenge is to understand how this statistical physics picture yields rubber elasticity. Treloar assumed in [77] that for a piece of rubber undergoing some macroscopic deformation, the cross-links do not fluctuate and follow the macroscopic deformation, whereas between two cross-links, the chains fluctuate. This is the so-called affine assumption. Treloar's model is in rather good agreement with mechanical experiments in small deformation. In large deformation however, it overestimates the stress. A natural possibility to relax Treloar's model consists in relaxing the affine assumption while keeping the network description, which allows one to distinguish between different rubbers. This can be done by assuming that the deformation of the cross-links minimizes the free energy of the polymer chains, the deformation being fixed at the boundary of the macroscopic domain D. This gives rise to a "variational model". The analysis of the asymptotic behavior of this model as the typical length of a polymer chain vanishes has the same flavor as the homogenization theory of integral functionals in nonlinear elasticity (see [55], [73] in the periodic setting, and [56] in the random setting).

Our aim is to relate qualitatively and quantitatively the (precise but unpractical) statistical physics picture to explicit macroscopic constitutive laws that can be used for practical purposes.

In collaboration with R. Alicandro (Univ. Cassino, Italy) and M. Cicalese (Univ. Munich, Germany), A. Gloria analyzed in [1] the (asymptotic) Γ-convergence of the variational model for rubber, in the case when the polymer chain network is represented by some ergodic random graph. The easiest such graph is the Delaunay tessellation of a point set generated as follows: random hard spheres of some given radius ρ are picked randomly until the domain is jammed (the so-called random parking measure of intensity ρ). With M. Penrose (Univ. Bath, UK), A. Gloria studied this random graph in this framework [5]. With P. Le Tallec (Mechanics department, Ecole polytechnique, France), M. Vidrascu (project-team REO, Inria Paris-Rocquencourt), and A. Gloria introduced and tested in [65] a numerical algorithm to approximate the homogenized energy density, and observed that this model compares well to rubber elasticity qualitatively.

These preliminary results show that the variational model has the potential to explain qualitatively and quantitatively how rubber elasticity emerges from polymer physics. In order to go further and obtain more quantitative results and rigorously justify the model, we have to address several questions of analysis, modelling, scientific computing, inverse problems, and physics.

3.2. Quantitative stochastic homogenization

Whereas the approximation of homogenized coefficients is an easy task in periodic homogenization, this is a highly nontrivial task for stochastic coefficients. This is in order to analyze numerical approximation methods of the homogenized coefficients that F. Otto (MPI for mathematics in the sciences, Leipzig, Germany) and A. Gloria obtained the first quantitative results in stochastic homogenization [3]. The development of a complete stochastic homogenization theory seems to be ripe for the analysis and constitutes the second major objective of this section.

In order to develop a quantitative theory of stochastic homogenization, one needs to quantitatively understand the corrector equation (3). Provided A is stationary and ergodic, it is known that there exists a unique random field ϕ_ξ which is a distributional solution of (3) almost surely, such that $\nabla\phi_\xi$ is a stationary random field with bounded second moment $\left\langle|\nabla\phi_\xi|^2\right\rangle < \infty$, and with $\phi(0) = 0$. Soft arguments do not allow to prove that ϕ_ξ may be chosen stationary (this is wrong in dimension $d = 1$). In [3], [4] F. Otto and A. Gloria proved that, in the case of discrete elliptic equations with iid conductances, there exists a unique stationary corrector ϕ_ξ with vanishing expectation in dimension $d > 2$. Although it cannot be bounded, it has bounded finite moments of any order:

$$\langle|\phi_\xi|^q\rangle < \infty \text{ for all } q \geq 1. \tag{6}$$

They also proved that the variance of spatial averages of the energy density $(\xi + \nabla\phi_\xi) \cdot A(\xi + \nabla\phi_\xi)$ on balls of radius R decays at the rate R^{-d} of the central limit theorem. These are the *first optimal quantitative results* in stochastic homogenization.

The proof of these results, which is inspired by [74], is based on the insight that coefficients such as the Poisson random inclusions are special in the sense that the associated probability measure satisfies a spectral gap estimate. Combined with elliptic regularity theory, this spectral gap estimate quantifies ergodicity in stochastic homogenization. This systematic use of tools from statistical physics has opened the way to the quantitative study of stochastic homogenization problems, which we plan to fully develop.

3.3. Nonlinear Schrödinger equations

As well known, the (non)linear Schrödinger equation

$$\partial_t\varphi(t,x) = -\Delta\varphi(t,x) + \lambda V(x)\varphi(t,x) + g|\varphi|^2\varphi(t,x), \quad \varphi(0,x) = \varphi_0(x) \tag{7}$$

with coupling constants $g \in \mathbb{R}, \lambda \in \mathbb{R}_+$ and real potential V (possibly depending also on time) models many phenomena of physics.

When in the equation (5) above one sets $\lambda = 0, g \neq 0$, one obtains the nonlinear (focusing of defocusing) Schrödinger equation. It is used to model light propagation in optical fibers. In fact, it then takes the following form:

$$i\partial_z\varphi(t,z) = -\beta(z)\partial_t^2\varphi(t,z) + \gamma(z)|\varphi(t,z)|^2\varphi(z,t), \tag{8}$$

where β and γ are functions that characterize the physical properties of the fiber, t is time and z the position along the fiber. Several issues are of importance here. Two that will be investigated within the MEPHYSTO project are: the influence of a periodic modulation of the fiber parameters β and γ and the generation of so-called “rogue waves” (which are solutions of unusually high amplitude) in such systems.

If $g = 0, \lambda \neq 0$, V is a random potential, and φ_0 is deterministic, this is the standard random Schrödinger equation describing for example the motion of an electron in a random medium. The main issue in this setting is the determination of the regime of Anderson localization, a property characterized by the boundedness in time of the second moment $\int x^2 |\varphi(t,x)|^2 dx$ of the solution. If this second moment remains bounded in time, the solution is said to be localized. Whereas it is known that the solution is localized in one dimension for all (suitable) initial data, both localized and delocalized solutions exist in dimension 3 and it remains a major open problem today to prove this, cf. [61].

If now $g \neq 0, \lambda \neq 0$ and V is still random, but $|g| \ll \lambda$, a natural question is whether, and in which regime, one-dimensional Anderson localization perdures. Indeed, Anderson localization can be affected by the presence of the nonlinearity, which corresponds to an interaction between the electrons or atoms. Much numerical and some analytical work has been done on this issue (see for example [64] for a recent work at PhLAM, Laser physics department, Univ. Lille 1), but many questions remain, notably on the dependence of the result on the initial conditions, which, in a nonlinear system, may be very complex. The cold atoms team of PhLAM (Garreau-Szriftgiser) is currently setting up an experiment to analyze the effect of the interactions in a Bose-Einstein condensate on a closely related localization phenomenon called "dynamical localization", in the kicked rotor, see below.

3.4. Processes in random environment

In the course of developing a quantitative theory of stochastic homogenization of discrete elliptic equations, we have introduced new tools to quantify ergodicity in partial differential equations. These tools are however not limited to PDEs, and could also have an impact in other fields where an evolution takes place in a (possibly dynamic) random environment and an averaging process occurs. The goal is then to understand the asymptotics of the motion of the particle/process.

For a random walker in a random environment, the Kipnis-Varadhan theorem ensures that the expected squared-position of the random walker after time t is of order t (the prefactor depends on the homogenized coefficients). If instead of a random walk among random conductances we consider a particle with some initial velocity evolving in a random *potential* field according to the Newton law, the averaged squared-position at time t is expected to follow the scaling law t^2, see [44]. This is called stochastic acceleration.

Similar questions arise when the medium is reactive (that is, when the potential is modified by the particle itself). The approach to equilibrium in such systems was observed numerically and explained theoretically, but not completely proven, in [58].

Another related and more general direction of research is the validity of *universality principle* of statistical physics, which states that the qualitative behaviour of physical systems depend on the microscopic details of the system only through some large-scale variables (the thermodynamic variables). Therefore, it is a natural problem in the field of interacting particle systems to obtain the macroscopic laws of the relevant thermodynamical quantities, using an underlying microscopic dynamics, namely particles that move according to some prescribed stochastic law. Probabilistically speaking, these systems are continuous time Markov processes.

4. Application Domains

4.1. Mechanics of heterogeneous media

The mechanics of heterogeneous materials aims at characterizing the macroscopic properties of heterogeneous materials using the properties of their constituents.

The homogenization theory is a natural tool for this task. In particular, for linear problems (linear conductivity or linear elasticity), the macroscopic properties are encoded into a single (conductivity or elasticity) homogenized tensor. The numerical approximation of this homogenized tensor is a typical objective of quantitative homogenization.

For nonlinear problems, such as rubber elasticity, the macroscopic properties are no longer characterized by a single tensor, but rather by a nonlinear energy density. Our aim is to relate qualitatively and quantitatively the (precise but unpractical) statistical physics picture to explicit macroscopic constitutive laws that can be used for practical purposes. This endeavor is relevant both in science and technology. The rigorous derivation of rubber elasticity from polymer-physics was indeed emphasized by John Ball as an important open problem of nonlinear elasticity in his survey [50] on the field. Its solution could shed light on some aspects of polymer-physics. The associated ab initio derivation of constitutive laws (as an alternative to phenomenological laws) would also be of interest to computational mechanics and rubber industry.

For this application domain, we work in close collaboration with physicists (FranÃ§ois Lequeux, ESPCI) and researchers from mechanics and computational mechanics (Patrick Le Tallec, Ecole polytechnique).

4.2. Numerical simulation in heterogeneous media

Solving numerically PDEs in highly heterogeneous media is a problem encountered in many situations, such as the transport of pollutants or the design of oil extraction strategies in geological undergrounds. When such problems are discretized by standard numerical methods the number of degrees of freedom may become prohibitive in practice, whence the need for other strategies.

Numerical solution methods inspired by asymptotic analysis are among the very few feasible alternatives, and started fifteen years ago with the contributions of Hou and Wu [68] , Arbogast [47] etc. We refer to [62], [78],[2] for a recent state of the art. Numerical homogenization methods usually amount to looking for the solution of the problem (1) in the form $u_\varepsilon(x) \simeq u_0(x) + \varepsilon \nabla u_0(x) \cdot \Phi(x, \frac{x}{\varepsilon})$, where Φ is a proxy for the corrector field computed locally at point x (in particular, one does not use explicitly that the problem is periodic so that the method can be used for more general coefficients) and u_0 is a function which does not oscillate at scale.

Relying on our quantitative insight in stochastic homogenization, a first task consists in addressing the three following prototypical academic examples: periodic, quasi-periodic, and stationary ergodic coefficients with short range dependence, cf. [25]. The more ambitious challenge is to address more complex coefficients (of interest to practioners), and design adaptive and efficient algorithms for diffusion in heterogeneous media.

4.3. Laser physics

Our contribution to the analysis of models in laser physics is motivated by the LabEx CEMPI (Centre Européen pour les Mathématiques, la Physique et leurs Interactions, a large eight-year research and training project approved by the French government in February 2012 as a "Laboratoire d'Excellence" and an initiative of mathematicians and physicists of the Université Lille 1). For this application domain, we work in close collaboration with physicists, which ensures our direct impact on these scientific issues. We focus on two applications: optical fibers and cold atoms.

In collaboration with physicists from the PhLAM laboratory in Lille, we aim at developping new techniques for the numerical integration of a family of 1D Schrödinger-like equations modelling the propagation of laser pulses in optical fibers. The questions arising are challenging since physicists would like to have fairly fast and cheap methods for their problems, with correct qualitative and quantitative behaviors. Another point is that they are interested in methods and codes that are able to handle different physical situations, hence different terms in the NLS equation. To meet these requirements, we will have to use numerical time-integration techniques such as splitting methods or exponential Runge-Kutta methods, space discretization techniques such as finite differences and fast Fourier transforms, and absorbent boundary conditions. Our goal, together with the physicists is to be able to reproduce numerically the results of the experiments they make in actual optical fibers, and then to be able to tune parameters numerically to get more insight into the appearance of rogue waves beyond the dispersive blowup phenomenon.

Recall that the Schrödinger equation also describes Bose-Einstein condensates. A second experimental team at PhLAM projects to study questions related to Anderson localization in such condensates. In fact, they will realize the "kicked rotor" (see [60]), which provides a paradigm for Anderson localization, in a Bose-Einstein condensate. We plan to collaborate with them on the theoretical underpinnings of their findings, which pose many challenging questions.

5. Highlights of the Year

5.1. Highlights of the Year

The team obtained two striking results in 2016:

- In collaboration with O. Blondel, T. Franco, and P. Gonçalves, M. Simon has made significant progress towards the *weak KPZ universality conjecture*, which states that a large class of one-dimensional weakly asymmetric conservative systems should converge to the KPZ equation, cf. [28], [7].
- In collaboration with F. Otto, M. Duerinckx and A. Gloria developed a complete theory of fluctuations in stochastic homogenization, cf. [39].

6. New Software and Platforms

6.1. MODULEF

FUNCTIONAL DESCRIPTION

The numerical method to approximate the constitutive laws for rubber elasticity derived from polymer physics are implemented in the Inria software Modulef.

It is based on : - algorithms from stochastic geometry to generate suitable polymer networks, - Delaunay tessellation algorithms to deal with steric effects (courtesy of the Inria project-team GAMMA2), - the introduction of 1-dimensional finite elements for the polymer-chains in Modulef.

- Participants: Marina Vidrascu and Antoine Gloria
- Contact: Marina Vidrascu
- URL: https://www.rocq.inria.fr/modulef/

7. New Results

7.1. Macroscopic behaviors of large interacting particle systems

7.1.1. Stochastic acceleration and approach to equilibrium

S. De Bièvre, Carlos Mejia-Monasterio (Madrid) and Paul E. Parris (Missouri) [57] studied thermal equilibration in a two-component Lorentz gas, in which the obstacles are modeled by rotating disks. They show that a mechanism of dynamical friction leads to a fluctuation-dissipation relation that is responsible for driving the system to equilibrium.

Stephan De Bièvre, Jeremy Faupin (Metz) and Schuble (Metz) [59] studied a related model quantum mechanically. Here a quantum particle moves through a field of quantized bose fields, modeling membranes that exchange energy and momentum with the particle. They establish a number of spectral properties of this model, that will be essential to study the time-asymptotic behavior of the system.

S. De Bièvre and collaborators analyse in [20] a multi-particle, kinetic version of a Hamiltonian model describing the interaction of a gas of particles with a vibrating medium. They prove existence results for weak solutions, and identify an asymptotic regime where the model, quite surprisingly, approaches the attractive Vlasov—Poisson system.

7.1.2. *Towards the weak KPZ universality conjecture*

One may start by considering the microscopic system in equilibrium (its measure is parametrized by the thermodynamical quantities under investigation). By removing the mean to the empirical measure and by scaling it properly, one would like to show that the random process, obtained by this rescaling, converges, as the size of the system is taken to infinity, to another random process which is a solution of some generalized stochastic PDE. Thanks to the remarkable recent result of M. Jara and P. Gonçalves [66], one has now all in hands to establish the latter result for a particular stochastic PDE known as the stochastic Burgers equation, and its companion, the Kardar-Parisi-Zhang (KPZ) equation. Indeed, in the latter paper, the authors introduce a new tool, called the second order Boltzmann-Gibbs principle, which permits to replace certain additive functionals of the dynamics by similar functionals given in terms of the density of the particles.

In [28], M. Simon in collaboration with T. Franco and P. Gonçalves, investigate the case of a microscopic dynamics with local defects, which is much harder. More precisely, the microscopic particle system is locally perturbed, and depending on the type of perturbation, the macroscopic laws can hold different boundary conditions. Since the ideas of [66] do not apply to the model considered there, they propose a new way to estimate the error in the replacement performed in the Boltzmann-Gibbs principle.

In the same spirit, M. Simon in collaboration with O. Blondel and P. Gonçalves investigate in [7] the class of kinetically constrained lattice gases that have been introduced and intensively studied in the literature in the past few years. In these models, particles are subject to restrictive constraints that make both approaches of [66] and [28] not work, so that new mathematical tools are needed. The main technical difficulty is that their model exhibits configurations that do not evolve under the dynamics and are locally non-ergodic. Their proof does not impose any knowledge on the spectral gap for the microscopic models. Instead, it relies on the fact that, under the equilibrium measure, the probability to find a blocked configuration in a finite box is exponentially small in the size of the box.

With these two recent results, M. Simon and coauthors contribute towards the *weak KPZ universality conjecture*, which states that a large class of one-dimensional weakly asymmetric conservative systems should converge to the KPZ equation.

7.1.3. *Diffusion and fractional diffusion of energy*

The rigorous derivation of the heat equation from deterministic systems of Newtonian particles is one of the most fundamental questions in mathematical physics. The main issue is that the existence of conservation laws and the high number of degrees of freedom impose very poor ergodic properties to the associated dynamical systems. A possible way out of this lack of ergodicity is to introduce stochastic models, in such a way that in one hand ergodicity issues are solved by the stochastic dynamics and in the other hand the qualitative behaviour of the system is not modified by the randomness. In these models, one starts with a chain of oscillators with a Hamiltonian dynamics, and one adds a stochastic component in such a way that the fundamental conservation laws (energy, momentum and *stretch* in this case) are maintained, and the corresponding Gibbs measures become ergodic.

It was already proved in [51] that these stochastic chains model correctly the behaviour of the conductivity. In particular, it is prove that Fourier law holds in dimension $d \geq 3$ if energy and momentum are conserved, and in any dimension if only energy is conserved. Once the conductivity has been successfully understood, one investigates the existence of the *hydrodynamic limit*, which fully describes the macroscopic evolution of the *empirical profiles* associated to the conserved quantity. In [41], M. Simon in collaboration with T. Komorowski and S. Olla consider the unpinned harmonic chain where the velocities of particles can randomly change sign. The only conserved quantities of the dynamics are the energy and the elongation. Using a diffusive space-time scaling, the profile of elongation evolves independently of the energy and follows a linear diffusive equation.

The energy profile evolves following a non-linear diffusive equation involving the elongation. The presence of non-linearity makes the macroscopic limit non-trivial, and its mathematical proof requires very sophisticated arguments.

In [52] and [69] it has been previously shown that in the case of one-dimensional harmonic oscillators with noise that preserves the momentum, the scaling limit of the energy fluctuations is ruled by the *fractional* heat equation

$$\partial_t u = -(-\Delta)^{3/4} u.$$

This equation does not only predict the superdiffusivity of energy in momentum-conserving models, but it also predicts the speed at which it diverges. This result opens a way to a myriad of open problems. The main goal is to observe anomalous fractional superdiffusion type limit in the context of low dimensional asymmetric systems with several conserved quantities. In two recent papers by M. Simon in collaboration with C. Bernardin, P. Gonçalves, M. Jara, M. Sasada [53] & [32], they confirmed rigorously recent Spohn's predictions on the Lévy form of the energy fluctuations for a harmonic chain perturbed by an energy-volume conservative noise. In [32] they also showed the existence of a crossover between a normal diffusion regime and a fractional superdiffusion regime by tuning a parameter of a supplementary stochastic noise conserving the energy but not the volume.

7.2. Qualitative results in homogenization

7.2.1. Isotropy and loss of ellipticity in periodic homogenization

Since the seminal contribution of Geymonat, Müller, and Triantafyllidis, it is known that strong ellipticity is not necessarily conserved by homogenization in linear elasticity. This phenomenon is typically related to microscopic buckling of the composite material. In [24] G. Francfort and A. Gloria study the interplay between isotropy and strong ellipticity in the framework of periodic homogenization in linear elasticity. Mixtures of two isotropic phases may indeed lead to loss of strong ellipticity when arranged in a laminate manner. They show that if a matrix/inclusion type mixture of isotropic phases produces macroscopic isotropy, then strong ellipticity cannot be lost.

7.2.2. From polymer physics to nonlinear elasticity

OIn [23], M. Duerinckx and A. Gloria succeeded in relaxing one of the two unphysical assumptions made in [1] on the growth of the energy of polymer chains. In particular, [23] deals with the case when the energy of the polymer chain is allowed to blow up at finite deformation.

7.2.3. The Clausius-Mossotti formula

In the mid-nineteenth century, Clausis, Mossotti and Maxwell essentially gave a first order Taylor expansion for (what is now understood as) the homogenized coefficients associated with a constant background medium perturbed by diluted spherical inclusions. Such an approach was recently used and extended by the team MATHERIALS to reduce the variance in numerical approximations of the homogenized coefficients, cf. [46], [45], [72]. In [22], M. Duerinckx and A. Gloria gave the first rigorous proof of the Clausius-Mossotti formula and provided the theoretical background to analyze the methods introduced in [72].

7.3. Quantitative results in stochastic homogenization

7.3.1. Quantitative results for almost periodic coefficients

In [6], S. Armstrong, A. Gloria and T. Kuusi (Aalto University) obtained the first improvement over the thirty year-old result by Kozlov [70] on almost periodic homogenization. In particular they introduced a class of almost periodic coefficients which are not quasi-periodic (and thus strictly contains the Kozlov class) and for which almost periodic correctors exist. Their approach combines the regularity theory developed by S. Armstrong and C. Smart in [49] and adapted to the almost periodic setting by S. Armstrong and Z. Shen [48], a new quantification of almost-periodicity, and a sensitivity calculus in the spirit of [3].

7.3.2. Optimal stochastic integrability in stochastic homogenization

In [40] A. Gloria and F. Otto consider uniformly elliptic coefficient fields that are randomly distributed according to a stationary ensemble of a finite range of dependence. They show that the gradient and flux $(\nabla\phi, a(\nabla\phi + e))$ of the corrector ϕ, when spatially averaged over a scale $R \gg 1$ decay like the CLT scaling $R^{-d/2}$. They establish this optimal rate on the level of *sub-Gaussian* bounds in terms of the stochastic integrability, and also establish a suboptimal rate on the level of optimal Gaussian bounds in terms of the stochastic integrability. The proof unravels and exploits the self-averaging property of the associated semi-group, which provides a natural and convenient disintegration of scales, and culminates in a propagator estimate with strong stochastic integrability. As an application, they characterize the fluctuations of the homogenization commutator, and prove sharp bounds on the spatial growth of the corrector, a quantitative two-scale expansion, and several other estimates of interest in homogenization.

7.3.3. A theory of fluctuations in stochastic homogenization

In [39], M. Duerinckx, A. Gloria, and F. Otto establish a path-wise theory of fluctuations in stochastic homogenization of linear elliptic equations in divergence form. More precisely they consider the model problem of a discrete equation with independent and identically distributed conductances (as considered in [27]). They identify a single quantity, which they call the homogenization commutator, that drives the fluctuations in stochastic homogenization in the following sense. On the one hand, this tensor-valued stationary random field satisfies a functional central limit theorem, and (when suitably rescaled) converges to a Gaussian white noise. On the other hand, the fluctuations of the gradient of the corrector, the fluctuations of the flux of the corrector, and the fluctuations of any solution of the PDE with random coefficients and localized right-hand side are characterized at leading order by the fluctuations of this homogenization commutator in a path-wise sense. As a consequence, when properly rescaled, the solution satisfies a functional central limit theorem, the gradient of the corrector converges to the Helmholtz projection of a Gaussian white noise, and the flux of the corrector converges to the Leray projection of the same white noise. Compared to previous contributions, our approach, based on the homogenization commutator, unravels the complete structure of fluctuations. It holds in any dimension $d \geq 2$, yields the first path-wise results, quantifies the limit theorems in Wasserstein distance, and only relies on arguments that extend to the continuum setting and to the case of systems.

7.4. Numerical methods for evolution equations

In [36] G. Dujardin analyzes an exponential integrator applied to the nonlinear Schrödinger equation with white noise dispersion. This models appears in optic fibers. Together with his co-author, he proves that this explicit scheme applied to the sctochastic PDE is of mean-square order 1. He uses it to illustrate a conjecture on the well-posedness of the equation in some regimes of the nonlinearity. Comparisons with several other schemes of the litterature are proposed. A last, another new (implicit) exponential integrators is proposed, which preserves the L^2-norm of the solution and is compared with the explicit one introduced beforehand.

7.5. Schrödinger equations

7.5.1. Nonlinear optical fibers

S. Rota Nodari, G. Dujardin, S. De Bièvre and collaborators continued their previous work on periodically modulated optical fibers with the experimental physicists of PhLAM [19]. They show that the nonlinear stage of modulational instability induced by parametric driving in the *defocusing* nonlinear Schrödinger equation can be accurately described by combining mode truncation and averaging methods, valid in the strong driving regime. The resulting integrable oscillator reveals a complex hidden heteroclinic structure of the instability. A remarkable consequence, validated by the numerical integration of the original model, is the existence of breather solutions separating different Fermi-Pasta-Ulam recurrent regimes.

In [42] S. de Bièvre and G. Dujardin analyze the formation of the Kuznetsov-Ma soliton of the 1D Schrödinger equation in the presence of periodic modulation satisfying an integrability condition. They show that this particular soliton has several compression points, the number, position and shape of which are controlled by the amplitude and the frequency of the modulation. They analyze the interplay between the frenquency of the soliton and the frequency of the modulation. Moreover, they show that one can suppress any component of the output spectrum of the soliton by a suitable choice of the amplitude and frequency of the modulation.

These works are part of the activities developped in the LabEx CEMPI.

7.5.2. Nonlinear Schrödinger equations

In [54], D. Bonheure, J.-B. Casteras and R. Nascimento obtained new results on the existence and qualitative properties of waveguides for a mixed-diffusion NLS. In particular, they proved the first existence results for waveguides with fixed mass and provided several qualitative descriptions of these.

S. De Bièvre and S. Rota Nodari continued their work on orbital stability of relative equilibria of Hamiltonian dynamical systems on Banach spaces, with a second paper [37], dealing with the situation where multi-dimensional invariance groups are present in the systems considered. They present a generalization of the Vakhitov-Kolokolov slope condition to this higher dimensional setting, and show how it allows to prove the local coercivity of the Lyapunov function, which in turn implies orbital stability. The method is applied to study the orbital stability of the plane waves of a system of two coupled nonlinear Schrödinger equations. They provide a comparison of their approach to the classical one by Grillakis-Shatah-Strauss.

7.6. Miscellaneous results

In [21] Mitia Duerinckx establishes the global well-posedness of a family of equations, which are obtained in certain regimes — in a joint work in preparation with Sylvia Serfaty — as the mean-field evolution of the supercurrent density in a (2D section of a) type-II superconductor with pinning and with imposed electric current. General vortex-sheet initial data are also considered, and the uniqueness and regularity properties of the solution are investigated.

In [33], [8], [11], [12], D. Bonheure, J.-B. Casteras and collaborators made bifurcation analysis and constructed multi-layer solutions of the Lin-Ni-Takagi and Keller-Segel equations, which come from the Keller-Segel system of chemotaxis in specific cases. A remarkable feature of the results is that the layers do not accumulate to the boundary of the domain but satisfy an optimal partition problem contrary to the previous type of solutions constructed for these models.

In [16], [17], [35], J.-B. Casteras and collaborators study different problems related to the existence of constant mean curvature hypersurfaces with prescribed asymptotic boundary on Cartan-Hadamard manifold. In particular, they obtained the first existence results for minimal graphs with prescribed asymptotic Dirichlet data under a pointwise pinching condition for sectionals curvatures.

S. De Bièvre and co-workers present in [67] a general approach to calculating the entanglement of formation for superpositions of two-mode coherent states, placed equidistantly on a circle in phase space. In the particular case of rotationally-invariant circular states the value of their entanglement is shown to be given by analytical expressions. They analyse the dependence of the entanglement on the radius of the circle and number of components in the superposition.

A. Benoit continues his analysis of hyperbolic equations in corner spaces. He addresses in [30] the rigorous construction of geometric optics expansions for weakly well-posed hyperbolic corner problems. He studies in [31] the semi-group stability for finite difference discretizations of hyperbolic systems of equations in corner domains, extending previous results of Coulombel & Gloria and Coulombel in the case of the halfspace.

8. Partnerships and Cooperations

8.1. National Initiatives

8.1.1. ANR BECASIM

G. Dujardin is a member of the ANR BECASIM project (http://becasim.math.cnrs.fr/). This ANR project gathers mathematicians with theoretical and numerical backgrounds together with engineers. The objective is to develop numerical methods to accurately simulate the behavior of Bose-Einstein condensates.

Title: Simulation numérique avancée pour les condensats de Bose-Einstein.

Type: Modèles Numériques - 2012

ANR reference: ANR-12-MONU-0007

Coordinator: Ionut DANAILA, Université de Rouen.

Duration: January 2013 - December 2017.

Partners: Université Lille 1, UPMC, Ecole des Ponts ParisTech, Inria-Nancy Grand-Est, Université Montpellier 2, Université Toulouse 3.

8.1.2. ANR EDNHS

M. Simon is a member of the ANR EDNHS project.

Title: Diffusion de l'énergie dans des système hamiltoniens bruités.

Type: Défi de tous les savoirs (DS10) 2014.

ANR reference: ANR-14-CE25-0011.

Coordinator: Cédric Bernardin, Université de Rennes.

Duration: October 2014 - October 2019.

8.1.3. Labex CEMPI

Title: Centre Européen pour les Mathématiques, la Physique et leurs interactions

Coordinator: Stephan De Bièvre.

Duration: January 2012 - December 2019.

Partners: Laboratoire Paul Painlevé and Laser physics department (PhLAM), Université Lille 1.

The "Laboratoire d'Excellence" Centre Européen pour les Mathématiques, la Physique et leurs interactions (CEMPI), a project of the Laboratoire de Mathématiques Paul Painlevé and the Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM), was created in the context of the "Programme d'Investissements d'Avenir" in February 2012.

The association Painlevé-PhLAM creates in Lille a research unit for fundamental and applied research and for training and technological development that covers a wide spectrum of knowledge stretching from pure and applied mathematics to experimental and applied physics.

One of the three focus areas of CEMPI research is the interface between mathematics and physics. This focus area encompasses three themes. The first is concerned with key problems of a mathematical, physical and technological nature coming from the study of complex behaviour in cold atoms physics and non-linear optics, in particular fibre optics. The two other themes deal with fields of mathematics such as algebraic geometry, modular forms, operator algebras, harmonic analysis and quantum groups that have promising interactions with several branches of theoretical physics.

8.1.4. PEPS "Jeunes Chercheurs"

M. Simon obtained a CNRS grant "PEPS Jeunes Chercheurs" for a project in collaboration with Oriane Blondel (Université Lyon 1), Clément Erignoux (IMPA, Rio de Janeiro) and Makiko Sasada (Tokyo University)

8.1.5. MIS

Incentive Grant for Scientific Research (MIS) of the Fonds National de la Recherche Scientifique (Belgium)

Title: Patterns, Phase Transitions, 4NLS & BIon.

Coordinator: D. Bonheure.

Duration: January 2014 - December 2016.

Partner: Université libre de Bruxelles.

8.1.6. PDR

Research Project (PDR) of the Fonds National de la Recherche Scientifique (Belgium).

D. Bonheure is co-investigator of this PDR.

Title: Asymptotic properties of semilinear systems.

Coordinator: Christophe Troestler (UMons).

Duration: July 2014 - June 2018.

Partner: Université de Mons, Université catholique de Louvain, Université libre de Bruxelles.

8.2. European Initiatives

8.2.1. FP7 & H2020 Projects

8.2.1.1. QUANTHOM

Title: Quantitative methods in stochastic homogenization

Programm: FP7

Duration: February 2014 - January 2019

Coordinator: Université Libre de Bruxelles (Belgium)

Partner: Inria

Inria contact: A. Gloria

'This proposal deals with the development of quantitative tools in stochastic homogenization, and their applications to materials science. Three main challenges will be addressed. First, a complete quantitative theory of stochastic homogenization of linear elliptic equations will be developed starting from results I recently obtained on the subject combining tools originally introduced for statistical physics, such as spectral gap and logarithmic Sobolev inequalities, with elliptic regularity theory. The ultimate goal is to prove a central limit theorem for solutions to elliptic PDEs with random coefficients. The second challenge consists in developing an adaptive multiscale numerical method for diffusion in inhomogeneous media. Many powerful numerical methods were introduced in the last few years, and analyzed in the case of periodic coefficients. Relying on my recent results on quantitative stochastic homogenization, I have made a sharp numerical analysis of these methods, and introduced more efficient variants, so that the three academic examples of periodic, quasi-periodic, and random stationary diffusion coefficients can be dealt with efficiently. The emphasis of this challenge is put on the adaptivity with respect to the local structure of the diffusion coefficients, in order to deal with more complex examples of interest to practitioners. The last and larger objective is to make a rigorous connection between the continuum theory of nonlinear elastic materials and polymer-chain physics through stochastic homogenization of nonlinear problems and random graphs. Analytic and numerical preliminary results show the potential of this approach. I plan to derive explicit constitutive laws for rubber from polymer chain properties, using the insight of the first two challenges. This requires a good understanding of polymer physics in addition to qualitative and quantitative stochastic homogenization.'

8.2.2. Collaborations with Major European Organizations

Max Planck Institute for Mathematics in the Sciences (Germany)

Long-term collaboration with F. Otto on the development of a quantitative theory of stochastic homogenization of linear elliptic systems.

8.3. International Research Visitors

8.3.1. Visits of International Scientists

Milton Jara, Professor Adjunto, IMPA, Rio de Janeiro (Brazil), was an invited professor at Université Lille 1 funded by the LabeX CEMPI.

8.3.1.1. Internships

Pierre Mennuni, MA2 Université Lille 1, Internship, 3 months

8.3.1.2. Research Stays Abroad

M. Simon spent one month at Universidade Federal Fluminense (Niteroi, Brazil) in march 2016, sponsored by the "Réseau France-Brésil", as a guest of Freddy Hernandez.

S. De Bièvre visited C. Mejia-Monasterio at the Technical University of Madrid in June 2016.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. General Chair, Scientific Chair

D. Bonheure was General chair of the committee of the BRUXELLES-TORINO TALKS IN PDE'S MAY 2-5 2016 UNIVERSITÀ DEGLI STUDI DI TORINO DIPARTIMENTO DI MATEMATICA “GIUSEPPE PEANO”

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

D. Bonheure is associate editor at the Bulletin of the Belgian Mathematical Society - Simon Stevin (http://projecteuclid.org/info/euclid.bbms)

A. Gloria is associated editor at the North-Western European Journal of Mathematics (http://math.univ-lille1.fr/~nwejm/).

9.1.3. Invited Talks

D. Bonheure was invited speaker at

- 02/12/2016, Recent progress in Partial Differential Equations, Université Aix-Marseille
- 03/10/2016 au 07/10/2016, New trends in Partial Differential Equations, Centro De Giorgi, SNS Pisa
- 28/09/2016 au 30/09/2016, 3rd Conference on Recent Trends in Nonlinear Phenomena, Perugia
- 12/09/2016 au 17/09/2016, Partial Differential Equations and Related Topics – On the occasion of Giorgio Talenti's 75th birthday , Alghero

Stephan De Bièvre was invited speaker at

- Université Fourier, Grenoble, Mathematical Physics Seminar, April 2016
- Université de Nantes, Mathematical Physics Seminar, May 2016
- Spectral and Scattering theories in QFT, IV, Porquerolles, May 2016
- 11th AIMS Conference on Dynamical systems, ODE, and Applications, Orlando, July 2016
- Conference “Coherent States and their Applications: A Contemporary Panorama”, CIRM Marseille, November 2016

A. Gloria was invited speaker at

- ENS Rennes, January 2016
- GAMM conference, Paris, January 2016
- Winter school on stochastic homogenization, Augsburg, February 2016
- Bristish Mathematical Colloquium, Bristol, March 2016
- Courant Institute, NYU, New York, May 2016
- 15th European Mechanics of Materials Conference, Brussels, September 2016
- Probability seminar, Warwick, October 2016
- Workshop “Functional inequalities, heat kernels, and random processes”, Oberwolfach institute, December 2016

M. Simon was invited speaker at

- YEP XIII: Large Deviations for Interacting Particle Systems and Partial Differential Equations, Eindhoven (Netherlands)
- Workshop in Stochastic Analysis, Universidade Federal de Campinas, Brazil
- Tokyo University, Probability Seminar
- Workhsop "Large Scale Stochastic Dynamics", Oberwolfach institute, Germany

9.1.4. Leadership within the Scientific Community

D. Bonheure is member of the Executive board of the Belgian Mathematical Society.

S. De Bièvre is

- the scientific coordinator of the CEMPI (2012-2019),
- member of the drafting committee of the IDEX UDL, and of the delegation that presented the project to the jury in Paris (April 2015, January 2016),
- member of the Executive Committee of the International Association of Mathematical Physics (since 2012).

9.1.5. Scientific Expertise

D. Bonheure is a member of the ESF College of Expert Reviewers from 20 October 2016 to 19 October 2019

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

Licence : Denis Bonheure, Integral and differential calculus, 46h, L1 (mathematics & physics), Université Libre de Bruxelles, Belgium

Licence : Denis Bonheure, Mathematics for management engineering, 30h, L1 (mathematics & physics), Université Libre de Bruxelles, Belgium

Licence : Guillaume Dujardin, Integral and differential calculus, 60h, L2 (mathematics & physics), Université Libre de Bruxelles, Belgium

Licence : Stephan De Bièvre, Probability, 15h, L2 (physics), Université Lille 1

Licence : Stephan De Bièvre, Probability, 50h, L3 (mathematics), Université Lille 1

Licence : Stephan De Bièvre, Financial mathematics, 30h, L1 (economics), Université Lille 1

Master : Denis Bonheure, Variational methods and PDEs, 30h, M2, Université Libre de Bruxelles, Belgium

Master : Antoine Gloria, Anderson localization, 30h, M2, Université Libre de Bruxelles, Belgium

Master : Antoine Gloria, Fluctuations in stochastic homogenization, 4h30, Doctoral course, YEP XIII, Eindhoven, Netherlands

9.2.2. Supervision

PhD in progress : Mitia Duerinckx, Topics in stochastic homogenization of PDEs, 01/10/2014, A. Gloria & S. Serfaty (NYU).

PhD in progress : Pierre Mennuni, Université Lille 1, 01/10/2016, S. De Bièvre, A. De Laire (Lille 1) & G. Dujardin

PhD in progress : Hussein Cheikh-Ali, Université Libre de Bruxelles and Université de Lorraine, D. Bonheure (ULB) & F. Robert (Nancy)

PhD in progress : Robson Alves do Nascimento Filho, Université Libre de Bruxelles, D. Bonheure (ULB)

.

9.2.3. Juries

S. De Bièvre was in the jury of the PhD theses of A. Newman (Loughbourough, England), A. Vavasseur (Nice) and J.P. Miqueu (Rennes).

9.3. Popularization

M. Simon gave a talk at ESPE (Ecole Supérieure du Professorat et de l'Education) during the week "Semaine des mathématiques", about the mathematical properties of the official soccer ball. Marielle Simon is part of the program "MathenJeans".

10. Bibliography

Major publications by the team in recent years

[1] R. Alicandro, M. Cicalese, A. Gloria.*Integral representation results for energies defined on stochastic lattices and application to nonlinear elasticity*, in "Arch. Ration. Mech. Anal.", 2011, vol. 200, n$^{\text{o}}$ 3, p. 881–943.

[2] A. Gloria.*Numerical homogenization: survey, new results, and perspectives*, in "Esaim. Proc.", 2012, vol. 37, Mathematical and numerical approaches for multiscale problem.

[3] A. Gloria, F. Otto.*An optimal variance estimate in stochastic homogenization of discrete elliptic equations*, in "Ann. Probab.", 2011, vol. 39, n$^{\text{o}}$ 3, p. 779–856.

[4] A. Gloria, F. Otto.*An optimal error estimate in stochastic homogenization of discrete elliptic equations*, in "Ann. Appl. Probab.", 2012, vol. 22, n$^{\text{o}}$ 1, p. 1–28.

[5] A. Gloria, M. Penrose.*Random parking, Euclidean functionals, and rubber elasticity*, in "Comm. Math. Physics", 2013, vol. 321, n$^{\text{o}}$ 1, p. 1–31.

Publications of the year

Articles in International Peer-Reviewed Journal

[6] S. N. Armstrong, A. Gloria, T. Kuusi.*Bounded correctors in almost periodic homogenization*, in "Archive for Rational Mechanics and Analysis", 2016, vol. 222, n$^{\text{o}}$ 1, p. 393–426 [*DOI :* 10.1007/s00205-016-1004-0], https://hal.inria.fr/hal-01230991.

[7] O. BLONDEL, P. GONÇALVES, M. SIMON.*Convergence to the Stochastic Burgers Equation from a degenerate microscopic dynamics*, in "Electronic Journal of Probability", December 2016, vol. 21, n^{o} 69, 25 [*DOI :* 10.1214/16-EJP15], https://hal.archives-ouvertes.fr/hal-01295541.

[8] D. BONHEURE, J.-B. CASTERAS, B. NORIS.*Layered solutions with unbounded mass for the Keller–Segel equation*, in "Journal of Fixed Point Theory and Applications", 2016 [*DOI :* 10.1007/S11784-016-0364-2], https://hal.archives-ouvertes.fr/hal-01398930.

[9] D. BONHEURE, S. CINGOLANI, M. NYS.*Nonlinear Schrödinger equation: concentration on circles driven by an external magnetic field*, in "Calculus of Variations and Partial Differential Equations", 2016, vol. 55 [*DOI :* 10.1007/S00526-016-1013-8], https://hal.archives-ouvertes.fr/hal-01182834.

[10] D. BONHEURE, P. D 'AVENIA, A. POMPONIO.*On the electrostatic Born-Infeld equation with extended charges*, in "Communications in Mathematical Physics", 2016 [*DOI :* 10.1007/S00220-016-2586-Y], https://hal.archives-ouvertes.fr/hal-01182830.

[11] D. BONHEURE, M. GROSSI, B. NORIS, S. TERRACINI.*Multi-layer radial solutions for a supercritical Neumann problem*, in "Journal of Differential Equations", 2016, vol. 261 [*DOI :* 10.1016/J.JDE.2016.03.016], https://hal.archives-ouvertes.fr/hal-01182832.

[12] D. BONHEURE, C. GRUMIAU, C. TROESTLER.*Multiple radial positive solutions of semilinear elliptic problems with Neumann boundary conditions*, in "Nonlinear Analysis: Theory, Methods and Applications", 2016, vol. 147, p. 236-273 [*DOI :* 10.1016/J.NA.2016.09.010], https://hal.archives-ouvertes.fr/hal-01408548.

[13] D. BONHEURE, F. HAMEL.*One-dimensional symmetry and Liouville type results for the fourth order Allen-Cahn equation in R^N*, in "Chinese Annals of Mathematics - Series B", 2016, 25, https://hal.archives-ouvertes.fr/hal-01182688.

[14] D. BONHEURE, F. JURAJ, S. ALBERTO.*Qualitative properties of solutions to mixed-diffusion bistable equations*, in "Calculus of Variations and Partial Differential Equations", May 2016, vol. 55 [*DOI :* 10.1007/S00526-016-0987-6], https://hal.archives-ouvertes.fr/hal-01203710.

[15] D. BONHEURE, J. D. ROSSI, N. SAINTIER.*The limit as $p \to \infty$ in the eigenvalue problem for a system of p-Laplacians*, in "Annali di Matematica Pura ed Applicata", 2016, vol. 195, n^{o} 5, p. 1771–1785 [*DOI :* 10.1007/S10231-015-0547-2], https://hal.archives-ouvertes.fr/hal-01408551.

[16] J.-B. CASTERAS, E. HEINONEN, I. HOLOPAINEN.*Solvability of Minimal Graph Equation Under Pointwise Pinching Condition for Sectional Curvatures*, in "The Journal of Geometric Analysis", 2016 [*DOI :* 10.1007/S12220-016-9712-0], https://hal.archives-ouvertes.fr/hal-01398918.

[17] J.-B. CASTERAS, J. B. RIPOLL.*On asymptotic plateau's problem for CMC hypersurfaces on rank 1 symmetric spaces of noncompact type*, in "Asian Journal of Mathematics", 2016, vol. 20, n^{o} 4, p. 695 - 708 [*DOI :* 10.4310/AJM.2016.V20.N4.A5], https://hal.archives-ouvertes.fr/hal-01398916.

[18] C. CHAINAIS-HILLAIRET, T. GALLOUËT.*Study of a pseudo-stationary state for a corrosion model: existence and numerical approximation*, in "Nonlinear Analysis: Real World Applications", 2016, https://hal.inria.fr/hal-01147621.

[19] M. CONFORTI, A. MUSSOT, A. KUDLINSKI, S. ROTA NODARI, G. DUJARDIN, S. DE BIÈVRE, A. ARMAROLI, S. TRILLO.*Heteroclinic Structure of Parametric Resonance in the Nonlinear Schrödinger Equation*, in "Physical Review Letters", June 2016, vol. 117, n^{o} 1 [*DOI :* 10.1103/PHYSREVLETT.117.013901], https://hal.archives-ouvertes.fr/hal-01333882.

[20] S. DE BIÈVRE, A. VAVASSEUR, T. GOUDON.*Particles interacting with a vibrating medium: existence of solutions and convergence to the Vlasov–Poisson system*, in "SIAM Journal on Mathematical Analysis", 2016, vol. 48, n^{o} 6, p. 3984–4020, https://hal.archives-ouvertes.fr/hal-01286519.

[21] M. DUERINCKX.*Mean-field limits for some Riesz interaction gradient flows*, in "SIAM Journal on Mathematical Analysis", 2016, vol. 48, n^{o} 3, p. 2269-2300, https://hal.archives-ouvertes.fr/hal-01252661.

[22] M. DUERINCKX, A. GLORIA.*Analyticity of homogenized coefficients under Bernoulli perturbations and the Clausius-Mossotti formulas*, in "Archive for Rational Mechanics and Analysis", 2016, vol. 220, n^{o} 1, p. 297–361 [*DOI :* 10.1007/S00205-015-0933-3], https://hal.inria.fr/hal-01138797.

[23] M. DUERINCKX, A. GLORIA.*Stochastic homogenization of nonconvex unbounded integral functionals with convex growth*, in "Archive for Rational Mechanics and Analysis", 2016, vol. 221, n^{o} 3, p. 1511–1584 [*DOI :* 10.1007/S00205-016-0992-0], https://hal.inria.fr/hal-01192752.

[24] G. A. FRANCFORT, A. GLORIA.*Isotropy prohibits the loss of strong ellipticity through homogenization in linear elasticity*, in "Comptes Rendus Mathématique", 2016, vol. 354, p. 1139 - 1144 [*DOI :* 10.1016/J.CRMA.2016.09.014], https://hal.inria.fr/hal-01398518.

[25] A. GLORIA, Z. HABIBI.*Reduction of the resonance error in numerical homogenisation II: correctors and extrapolation*, in "Foundations of Computational Mathematics", 2016, vol. 16, n^{o} 1, p. 217–296 [*DOI :* 10.1007/S10208-015-9246-Z], https://hal.inria.fr/hal-00933234.

[26] A. GLORIA, D. MARAHRENS.*Annealed estimates on the Green functions and uncertainty quantification*, in "Annales de l'Institut Henri Poincaré (C) Non Linear Analysis", 2016, vol. 33, n^{o} 6, p. 1153–1197, 43 pages [*DOI :* 10.1016/J.ANIHPC.2015.04.001], https://hal.archives-ouvertes.fr/hal-01093386.

[27] A. GLORIA, J. NOLEN.*A quantitative central limit theorem for the effective conductance on the discrete torus*, in "Communications on Pure and Applied Mathematics", 2016, vol. 69, n^{o} 12, p. 2304–2348 [*DOI :* 10.1002/CPA.21614], https://hal.archives-ouvertes.fr/hal-01093352.

[28] M. SIMON, P. GONCALVES, T. FRANCO.*Crossover to the stochastic Burgers equation for the WASEP with a slow bond*, in "Communications in Mathematical Physics", 2016 [*DOI :* 10.1007/S00220-016-2607-X], https://hal.inria.fr/hal-01355447.

[29] M. SIMON, P. GONÇALVES, M. JARA.*Second order Boltzmann-Gibbs principle for polynomial functions and applications*, in "Journal of Statistical Physics", December 2016 [*DOI :* 10.1007/S10955-016-1686-6], https://hal.inria.fr/hal-01381009.

Other Publications

[30] A. BENOIT.*Geometric optics expansions for hyperbolic corner problems II : from weak stability to violent instability*, October 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01242899.

[31] A. BENOIT. *Semi-group stability of finite difference schemes in corner domains*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01417105.

[32] C. BERNARDIN, P. GONÇALVES, M. JARA, M. SIMON. *Interpolation process between standard diffusion and fractional diffusion*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01348503.

[33] D. BONHEURE, J.-B. CASTERAS, B. NORIS. *Multiple positive solutions of the stationary Keller-Segel system*, November 2016, 33 pages, https://hal.archives-ouvertes.fr/hal-01398922.

[34] D. BONHEURE, S. CINGOLANI, J. VAN SCHAFTINGEN. *The logarithmic Choquard equation : sharp asymptotics and nondegeneracy of the groundstate*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01414010.

[35] J.-B. CASTERAS, E. HEINONEN, I. HOLOPAINEN. *Dirichlet problem for f-minimal graphs*, November 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01398920.

[36] D. COHEN, G. DUJARDIN. *Exponential integrators for nonlinear Schrödinger equations with white noise dispersion*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01403036.

[37] S. DE BIEVRE, S. ROTA NODARI. *Orbital stability via the energy-momentum method: the case of higher dimensional symmetry groups*, May 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01312534.

[38] M. DUERINCKX. *Well-posedness for mean-field evolutions arising in superconductivity*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01389812.

[39] M. DUERINCKX, A. GLORIA, F. OTTO. *The structure of fluctuations in stochastic homogenization*, February 2016, working paper or preprint, https://hal.inria.fr/hal-01398515.

[40] A. GLORIA, F. OTTO. *The corrector in stochastic homogenization: optimal rates, stochastic integrability, and fluctuations*, May 2016, working paper or preprint, https://hal.inria.fr/hal-01230985.

[41] T. KOMOROWSKI, S. OLLA, M. SIMON. *Macroscopic evolution of mechanical and thermal energy in a harmonic chain with random flip of velocities*, August 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01358979.

[42] G. G. L. TIOFACK, S. COULIBALY, M. TAKI, S. DE BIÈVRE, G. DUJARDIN. *Periodic modulations controlling Kuznetsov-Ma soliton formation in nonlinear Schrödinger equations*, July 2016, working paper or preprint, https://hal.inria.fr/hal-01403028.

References in notes

[43] G. AGRAWAL. *Nonlinear fiber optics*, Academic Press, 2006.

[44] B. AGUER, S. DE BIÈVRE, P. LAFITTE, P. E. PARRIS. *Classical motion in force fields with short range correlations*, in "J. Stat. Phys.", 2010, vol. 138, n^{o} 4-5, p. 780 – 814.

[45] A. ANANTHARAMAN, C. LE BRIS.*A numerical approach related to defect-type theories for some weakly random problems in homogenization*, in "Multiscale Model. Simul.", 2011, vol. 9, n$^{\mathrm{o}}$ 2, p. 513–544, http://dx.doi.org/10.1137/10079639X.

[46] A. ANANTHARAMAN, C. LE BRIS.*Elements of mathematical foundations for numerical approaches for weakly random homogenization problems*, in "Commun. Comput. Phys.", 2012, vol. 11, n$^{\mathrm{o}}$ 4, p. 1103–1143, http://dx.doi.org/10.4208/cicp.030610.010411s.

[47] T. ARBOGAST.*Numerical subgrid upscaling of two-phase flow in porous media*, in "Numerical treatment of multiphase flows in porous media (Beijing, 1999)", Berlin, Lecture Notes in Phys., Springer, 2000, vol. 552, p. 35–49.

[48] S. N. ARMSTRONG, Z. SHEN.*Lipschitz estimates in almost-periodic homogenization*, in "Commun. Pure Appl. Mathematics", September 2015.

[49] S. N. ARMSTRONG, C. K. SMART.*Quantitative stochastic homogenization of convex integral functionals*, in "Ann. Scientifiques de l'ENS", 2015.

[50] J. M. BALL.*Some open problems in elasticity*, in "Geometry, mechanics, and dynamics", New York, Springer, 2002, p. 3–59.

[51] G. BASILE, C. BERNARDIN, S. OLLA.*Thermal conductivity for a momentum conservative model*, in "Comm. Math. Phys.", 2009, vol. 287, n$^{\mathrm{o}}$ 1, p. 67–98, http://dx.doi.org/10.1007/s00220-008-0662-7.

[52] C. BERNARDIN, P. GONÇALVES, M. JARA.*3/4-fractional superdiffusion in a system of harmonic oscillators perturbed by a conservative noise*, in "Arch. Ration. Mech. Anal.", 2016, vol. 220, n$^{\mathrm{o}}$ 2, p. 505–542, http://dx.doi.org/10.1007/s00205-015-0936-0.

[53] C. BERNARDIN, P. GONÇALVES, M. JARA, M. SASADA, M. SIMON.*From normal diffusion to superdiffusion of energy in the evanescent flip noise limit*, in "J. Stat. Phys.", 2015, vol. 159, n$^{\mathrm{o}}$ 6, p. 1327–1368, http://dx.doi.org/10.1007/s10955-015-1235-8.

[54] D. BONHEURE, J.-B. CASTERAS, R. NASCIMENTO.*Orbitally stable standing waves of a mixed dispersion nonlinear Schrödinger equation*, in "Preprint", 2016.

[55] A. BRAIDES.*Homogenization of some almost periodic functionals*, in "Rend. Accad. Naz. Sci. XL", 1985, vol. 103, p. 261–281.

[56] G. DAL MASO, L. MODICA.*Nonlinear stochastic homogenization and ergodic theory*, in "J. Reine Angew. Math.", 1986, vol. 368, p. 28–42.

[57] S. DE BIÈVRE, C. MEJIA-MONASTERIO, E. P. PARRIS.*Preprint*, 2016.

[58] S. DE BIÈVRE, P. PARRIS.*Equilibration, generalized equipartition and diffusion in dynamical Lorentz gases*, in "J. Stat. Phys.", 2011, vol. 142, p. 356–385.

[59] S. DE BIÈVRE, J. FAUPIN, B. SCHUBNEL.*Spectral analysis of a model for quantum friction*, December 2015, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01246914.

[60] S. DE BIÈVRE, G. FORNI.*Transport properties of kicked and quasiperiodic Hamiltonians*, in "J. Statist. Phys.", 2010, vol. 90, n^o 5-6, p. 1201–1223.

[61] M. DISERTORI, W. KIRSCH, A. KLEIN, F. KLOPP, V. RIVASSEAU.*Random Schrödinger operators*, Panoramas et Synthèses, Société Mathématique de France, Paris, 2008, n^o 25.

[62] Y. EFENDIEV, T. HOU.*Multiscale finite element methods*, Surveys and Tutorials in the Applied Mathematical Sciences, Springer, New York, 2009, vol. 4, Theory and applications.

[63] P. FLORY.*Statistical mechanics of chain molecules*, Interscience Publishers, New York, 1969.

[64] J.-C. GARREAU, B. VERMERSCH.*Spectral description of the dynamics of ultracold interacting bosons in disordered lattices*, in "New. J. Phys.", 2013, vol. 15, 045030.

[65] A. GLORIA, P. LE TALLEC, M. VIDRASCU.*Foundation, analysis, and numerical investigation of a variational network-based model for rubber*, in "Continuum Mech. Thermodyn.".

[66] P. GONÇALVES, M. JARA.*Nonlinear fluctuations of weakly asymmetric interacting particle systems*, in "Arch. Ration. Mech. Anal.", 2014, vol. 212, n^o 2, p. 597–644, http://dx.doi.org/10.1007/s00205-013-0693-x.

[67] D. HOROSHKO, S. DE BIÈVRE, M. I. KOLOBOV, G. PATERA.*Entanglement of quantum circular states of light*, in "Phys.Rev.A", June 2016 [*DOI :* 10.1103/PHYSREVA.93.062323], https://hal.inria.fr/hal-01408553.

[68] T. HOU, X. WU.*A Multiscale Finite Element Method for Elliptic Problems in Composite Materials and Porous Media*, in "J. Comput. Phys.", 1997, vol. 134, p. 169–189.

[69] M. JARA, T. KOMOROWSKI, S. OLLA.*Limit theorems for additive functionals of a Markov chain*, in "Ann. Appl. Probab.", 2009, vol. 19, n^o 6, p. 2270–2300, http://dx.doi.org/10.1214/09-AAP610.

[70] S. M. KOZLOV.*Averaging of differential operators with almost periodic rapidly oscillating coefficients*, in "Mat. Sb. (N.S.)", 1978, vol. 107(149), n^o 2, p. 199–217, 317.

[71] S. KOZLOV.*The averaging of random operators*, in "Mat. Sb. (N.S.)", 1979, vol. 109(151), n^o 2, p. 188–202, 327.

[72] F. LEGOLL, W. MINVIELLE.*A control variate approach based on a defect-type theory for variance reduction in stochastic homogenization*, in "Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal", 2015, vol. 13, n^o 2, p. 519-550, https://hal.archives-ouvertes.fr/hal-01053459.

[73] S. MÜLLER.*Homogenization of nonconvex integral functionals and cellular elastic materials*, in "Arch. Rat. Mech. Anal.", 1987, vol. 99, p. 189–212.

[74] A. NADDAF, T. SPENCER.*Estimates on the variance of some homogenization problems*, Preprint, 1998.

[75] G. PAPANICOLAOU, S. VARADHAN.*Boundary value problems with rapidly oscillating random coefficients*, in "Random fields, Vol. I, II (Esztergom, 1979)", Amsterdam, Colloq. Math. Soc. János Bolyai, North-Holland, 1981, vol. 27, p. 835–873.

[76] C. SULEM, P.-L. SULEM.*The nonlinear Schrödinger equation*, Springer-Verlag, New-York, 1999.

[77] L. TRELOAR.*The Physics of Rubber Elasticity*, Oxford at the Clarendon Press, Oxford, 1949.

[78] E. WEINAN.*Principles of multiscale modeling*, Cambridge University Press, Cambridge, 2011, xviii+466.

Project-Team MINT

Methods and tools for gestural interactions

IN PARTNERSHIP WITH:
CNRS

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Interaction and visualization

Table of contents

Project-Team MINT

Creation of the Team: 2010 January 01, updated into Project-Team: 2012 January 01, end of the Project-Team: 2016 December 31

Keywords:

Computer Science and Digital Science:

5.1.2. - Evaluation of interactive systems
5.1.3. - Haptic interfaces
5.1.5. - Body-based interfaces
5.6. - Virtual reality, augmented reality
5.7.2. - Music

Other Research Topics and Application Domains:

9.1. - Education
9.2. - Art
9.5.10. - Digital humanities

1. Members

Faculty Members

Laurent Grisoni [Team leader, Univ. Lille I, Professor, HDR]
Frederic Giraud [Univ. Lille I, Associate Professor, HDR]
Fabrice Aubert [Univ. Lille I, Associate Professor]
Francesco de Comité [Univ. Lille I, Associate Professor]
Betty Semail [Univ. Lille I, Professor, HDR]
Florent Berthaut [Univ. Lille III, Associate Professor, associate member]
Christophe Giraud-Audine [Arts & Métiers Paris tech, Associate Professor, associate member]

Technical Staff

Michel Amberg [Univ. Lille I]
Erwan Douaille [Inria]
David Demol [Inria, from Jun 2016 (MATRICE project)]

PhD Students

Cagan Arslan [Univ. Lille I]
Olivier Capra [Univ. Lille I, from Oct 2016]
Wael Ben Messaoud [Univ. Lille I]
Nicolas Bremard [Univ. Lille 1]
Sofiane Ghenna [Univ. Lille I]
Vincent Gouezou [CIFRE]
Farzan Kalantari [Univ. Lille I]
Charlotte Planckeel [Univ. Lille III]
Hanae Rateau [Univ. Lille I]
Eric Vezzoli [Univ. Lille I]
Ehsan Enferad [Univ. Lille I]

Post-Doctoral Fellows

Yosra Rekik [Univ. Lille I, fev.2016-dec. 2018]
David Gueorguiev [Inria, from Sep 2016]

Visiting Scientist
Orlando Erazo [Univ. Chile, PhD, from Nov 2015, for 3 months]

Administrative Assistant
Karine Lewandowski [Inria]

2. Overall Objectives

2.1. Overall Objectives

The Mint team focuses on *gestural interaction*, i.e. the use of gesture for human-computer interaction (HCI). The New Oxford American Dictionary defines *gesture* as *a movement of part of the body, especially a hand or the head, to express an idea or meaning*. In the particular context of HCI, we are more specifically interested in movements that a computing system can sense and respond to. A gesture can thus be seen as a function of time into a set of sensed dimensions that might include but are not limited to positional information (the pressure exerted on a contact surface being an example of non-positional dimension).

Simple pointing gestures have long been supported by interactive graphics systems and the advent of robust and affordable sensing technologies has somewhat broadened their use of gestures. Swiping, rotating and pinching gestures are now commonly supported on touch-sensitive devices, for example. Yet the expressive power of the available gestures remains limited. The increasing diversity and complexity of computer-supported activities calls for more powerful gestural interactions. Our goal is to foster the emergence of these new interactions, to further broaden the use of gesture by supporting more complex operations. We are developing the scientific and technical foundations required to facilitate the design, implementation and evaluation of these interactions. Our interests include:

- gestures captured using held, worn or touched objects or contactless perceptual technologies;
- transfer functions possibly used during the capture process;
- computational representations of the captured gestures;
- methods for characterising and recognising them;
- feedback mechanisms, and more particularly haptic ones;
- tools to facilitate the design and implementation of tactile and gestural interaction techniques;
- evaluation methods to assess the usability of these techniques.

3. Research Program

3.1. Human-Computer Interaction

The scientific approach that we follow considers user interfaces as means, not an end: our focus is not on interfaces, but on interaction considered as a phenomenon between a person and a computing system [26]. We *observe* this phenomenon in order to understand it, i.e. *describe* it and possibly *explain* it, and we look for ways to significantly *improve* it. HCI borrows its methods from various disciplines, including Computer Science, Psychology, Ethnography and Design. Participatory design methods can help determine users' problems and needs and generate new ideas, for example [30]. Rapid and iterative prototyping techniques allow to decide between alternative solutions [27]. Controlled studies based on experimental or quasi-experimental designs can then be used to evaluate the chosen solutions [32]. One of the main difficulties of HCI research is the doubly changing nature of the studied phenomenon: people can both adapt to the system and at the same time adapt it for their own specific purposes [29]. As these purposes are usually difficult to anticipate, we regularly *create* new versions of the systems we develop to take into account new theoretical and empirical knowledge. We also seek to *integrate* this knowledge in theoretical frameworks and software tools to disseminate it.

3.2. Numerical and algorithmic real-time gesture analysis

Whatever is the interface, user provides some curves, defined over time, to the application. The curves constitute a gesture (positional information, yet may also include pressure). Depending on the hardware input, such a gesture may be either continuous (e.g. data-glove), or not (e.g. multi-touch screens). User gesture can be multi-variate (several fingers captured at the same time, combined into a single gesture, possibly involving two hands, maybe more in the context of co-located collaboration), that we would like, at higher-level, to be structured in time from simple elements in order to create specific command combinations. One of the scientific foundations of the research project is an algorithmic and numerical study of gesture, which we classify into three points:

- *clustering*, that takes into account intrinsic structure of gesture (multi-finger/multi-hand/multi-user aspects), as a lower-level treatment for further use of gesture by application;
- *recognition*, that identifies some semantic from gesture, that can be further used for application control (as command input). We consider in this topic multi-finger gestures, two-handed gestures, gesture for collaboration, on which very few has been done so far to our knowledge. On the contrary, in the case of single gesture case (i.e. one single point moving over time in a continuous manner), numerous studies have been proposed in the current literature, and interestingly, are of interest in several communities: HMM [33], Dynamic Time Warping [35] are well-known methods for computer-vision community, and hand-writing recognition. In the computer graphics community, statistical classification using geometric descriptors has previously been used [31]; in the Human-Computer interaction community, some simple (and easy to implement) methods have been proposed, that provide a very good compromise between technical complexity and practical efficiency [34].
- *mapping to application*, that studies how to link gesture inputs to application. This ranges from transfer function that is classically involved in pointing tasks [28], to the question to know how to link gesture analysis and recognition to the algorithmic of application content, with specific reference examples.

We ground our activity on the topic of numerical algorithm, expertise that has been previously achieved by team members in the physical simulation community (within which we think that aspects such as elastic deformation energies evaluation, simulation of rigid bodies composed of unstructured particles, constraint-based animation... will bring up interesting and novel insights within HCI community).

3.3. Design and control of haptic devices

Our scientific approach in the design and control of haptic devices is focused on the interaction forces between the user and the device. We search of controlling them, as precisely as possible. This leads to different designs compared to other systems which control the deformation instead. The research is carried out in three steps:

- *identification:* we measure the forces which occur during the exploration of a real object, for example a surface for tactile purposes. We then analyse the record to deduce the key components – *on user's point of view* – of the interaction forces.
- *design:* we propose new designs of haptic devices, based on our knowledge of the key components of the interaction forces. For example, coupling tactile and kinesthetic feedback is a promising design to achieve a good simulation of actual surfaces. Our goal is to find designs which lead to compact systems, and which can stand close to a computer in a desktop environment.
- *control:* we have to supply the device with the good electrical signals to accurately output the good forces.

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Evita

EVITA is a tactile feedback tablet, produced by Hap2U SME company, based in grenoble. It is presented at CES in january 2017, the SME has been awarded a CES innovation award. This device is issued from a strong collaboration with MINT group. Eric Vezzoli PhD thesis, contributed significantly to this device. EVITA is a very generic interaction device, and several projects are currently being discussed for understanding the fields of applications of this device. It is also, in particular, the hardware support for our haptic book for children, described below, that is our second highlight for this raweb.

4.1.2. Haptic book

The first digital book augmented with a high fidelity feedback has been released in October 2016. Based on a scenario and illustrations made by Dominique Maes - an artist from Belgium - this haptic book was presented for the first time during "la nuit des bibliothèques" in Lille. The popularity of this project as well as its possible social outcomes were underlined in a paper in a national magazine ("Science et Avenir", November 2016)

4.1.3. Forum Oeuvres et Recherches

MINT played an active role in the "Oeuvres et recherches" project (http://www.cristal.univ-lille.fr/oeuvres-et-recherches/), a platform that aims at higlighting and supporting collaborations between researchers and artists in the Hauts-de-France and in Belgium. Since 2010, these collaborations have resulted in significant contributions for these two communities at the regional and national levels. Organised at the Université de Lille on December 2nd 2016, the F O O R event was an opportunity to review more than five years of art-science projects in the region and Belgium, highlighted more than 40 art-science projects, and more importantly to prepare the future and discuss strategies for supporting such projects.

4.1.4. ControllAR

The ControllAR project, started in 2016, investigates the appropriation of visual feedback on control surfaces for multimedia production systems. It has already yielded many results. The system and results of a study on electronic musicians were presented both as a paper and as a demo at the ACM Internation conference on Surfaces and Spaces (ISS 16) where it received a best demo award. The software was released and is available at http://forge.lifl.fr/ControllAR. ControllAR was also presented during multiple events, both for the general public and for electronic musicians. The project continues with the design of a portable hardware solution and a long term study of the effects of the system on musicians' playing techniques.

4.1.5. Awards

- Best demo award for *ControllAR : appropriation of visual feedback on control surfaces* [16] @ ACM International Conference on Interactive Surfaces and Spaces (ISS 16).
- Best work in progress at Eurohaptics 2016 for the work The human perception of transient frictional modulation, David Gueorguiev, Eric Vezzoli, André Mouraux, Betty Semail, Jean-Louis Thonnard
- SME Hap2U had a "CES innovation award", based on the collaboration that MINT group has with them (E-vita tactile feedback tablet) at CES (january 2017).

5. New Software and Platforms

5.1. ControllAR

FUNCTIONAL DESCRIPTION

ControllAR is a novel system that facilitates the appropriation of rich visual feedback on control surfaces through remixing of graphical user interfaces and augmented reality display.

- Contact: Florent Berthaut
- URL: http://forge.lifl.fr/ControllAR

5.2. GINA

- Participants: Nicolas Bremard and Laurent Grisoni
- Contact: Laurent Grisoni

5.3. Revil

FUNCTIONAL DESCRIPTION

Revil is an application for building and manipulating 3D SceneGraphs for Mixed-Reality Artistic Performances. It is based on OpenGL/GLSL(glfw, glm), OpenNI2, FLTK and is entirely controllable via OpenSoundControl messages. It relies on the approach of revealing virtual content in the physical space by intersecting it with performers and spectator's bodies and props.

- Contact: Florent Berthaut
- URL: http://forge.lifl.fr/Revil

5.4. TaxtelOSC

FUNCTIONAL DESCRIPTION

TaxtelOSC is a software wich uses the concept of taxtel to reproduce rich and dense tactile feedback on

- Contact: Laurent Grisoni
- URL: http://forge.lifl.fr/ControllAR

6. New Results

6.1. ControllAR: Appropriation of visual Feedback on Control Surfaces

Florent Berthaut, Alex Jones

Despite the development of touchscreens, many expert systems for working with digital multimedia content, such as in music composition and performance, video editing or visual performance, still rely on control surfaces. This can be due to the accuracy and appropriateness of their sensors, the haptic feedback that they offer, and most importantly the way they can be adapted to the specific subset of gestures and tasks that users need to perform. On the other hand, visual feedback on controllers remains limited and/or fixed, preventing similar personalizing. In this paper, we propose ControllAR, a novel system that facilitates the appropriation of rich visual feedback on control surfaces through remixing of graphical user interfaces and augmented reality display. We then use our system to study current and potential appropriation of visual feedback in the case of digital musical instruments and derive guidelines for designers and developers.

6.2. Talaria: Continuous Drag & Drop on a Wall Display

Hanaë Rateau, Yosra Rekik, Laurent Grisoni, Joaquim Jorge

We present an interaction technique combining tactile actions and Midair pointing to access out-of-reach content on large displays without the need to walk across the display. Users can start through a Touch gesture on the display surface and finish Midair by pointing to push content away or inversely to retrieve a content. The technique takes advantage of wellknown semantics of pointing in human-to-human interaction. These, coupled with the semantics of proximal relations and deictic proxemics make the proposed technique very powerful as it leverages on well-understood human-human interaction modalities. Experimental results show this technique to outperform direct tactile interaction on dragging tasks. From our experience we derive four guidelines for interaction with large-scale displays.

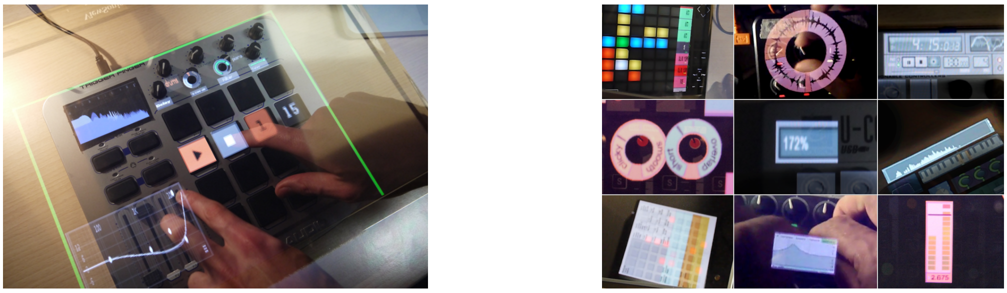


Figure 1. ControllAR: (left) ControllAR is used to augment a control surface with the remixed graphical user interface of music software, (right) Visual feedback designed by electronic musicians during our study belong to three categories: mappings feedback, processes feedback and content feedback.

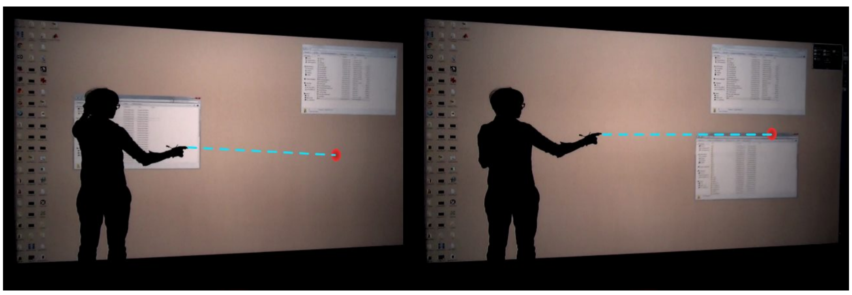
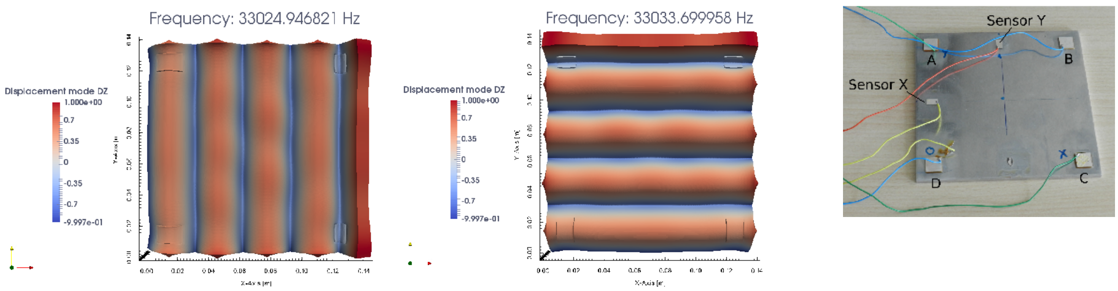


Figure 2. Talaria

6.3. Multi fngers interaction on a surface haptic display

Sofiane Ghenna, Christophe Giraud-Audine, Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail
In this study, we develop and implement a method for superimposing two vibration modes in order to produce different tactile stimuli on two fingers located in different positions. The tactile stimulation is based on the squeeze film effect which decreases the friction between a fingertip and a vibrating plate.

Experimental test have been conducted on a 1D tactile device. They show that it is possible to continuously control the friction on two fingers moving independently. Then, we developed the design of a 2D device based on the same principle, which gives rise to the design of a two fingers tactile display. Evaluations were conducted using a modal analysis with experimental validation.



(a) (b) (c)

Figure 3. Vibration modes and mode shapes using FEM with the position of the actuators (in white in a and b), and the prototype (c).

6.4. Finding the Minimum Perceivable Size of a Tactile Element on an Ultrasonic Based Haptic Tablet

Farzan Kalantari, Laurent Grisoni, Frédéric Giraud, Yosra Rekik
Tactile devices with ultrasonic vibrations (based on squeeze film effect) using piezoelectric actuators are one of the existing haptic feedback technologies. In this study we have performed two psychophysical experiments on an ultrasonic haptic tablet, in order to find the minimum size of a tactile element on which all the users are able to perfectly identify different types of textures. Our results show that the spatial resolution of the tactile element on haptic touchscreen actually varies, depending on the number and types of tactile feedback information. A first experiment exhibits three different tactile textures, chosen as being easily recognized by users. We use these textures in a second experiment, and evaluate minimal spatial area on which the chosen set of textures can be recognized. Among other, we find the minimal size depends on the texture nature.

6.5. BOEUF: A Unified Framework for Modeling and Designing Digital Orchestras

Florent Berthaut, Luke Dahl, Patricia Plénacoste

Orchestras of Digital Musical Instruments (DMIs) enable new musical collaboration possibilities, extending those of acoustic and electric orchestras. However the creation and development of these orchestras remain constrained. In fact, each new musical collaboration system or orchestra piece relies on a fixed number of musicians, a fixed set of instruments (often only one), and a fixed subset of possible modes of collaboration. In this paper, we describe a unified framework that enables the design of Digital Orchestras with potentially different DMIs and an expand-able set of collaboration modes. It relies on research done on analysis and classification of traditional and digital orchestras, on research in Collaborative Virtual Environments, and on interviews of musicians and composers. The BOEUF framework consists of a classification of modes of collaboration and a set of components for modelling digital orchestras. Integrating this framework into DMIs will enable advanced musical collaboration modes to be used in any digital orchestra, including spontaneous jam sessions.
Current work on this project consists in the implementation of BOEUF in the PureData programming language and in the study of its impact on musical collaboration during short improvised jam sessions.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

Hap2U SME is licenced two patents of MINT team.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. StimTac, 2015-2017

Participants: Frédéric Giraud [correspondant], Patricia Plénacoste, Laurent Grisoni, Michel Amberg, Nicolas Bremmard.

The aim of this project is to create the first digital book, enhanced with haptic feedback, in order to anticipate the integration of this technology into everyday products. This project adresses technological issues, like programming haptic content in a multimedia software, and design issues to understand how the haptic feedback is perceived by the users.

Stimtac is a book, and could thus be presented to non-specialists users and to a wide public during presentations, demos and foru. The scenario and the illustrations were made by Dominique Maes, a belgium artist, who did the digital book "bleu de toi" among other things. The Public Library of Lille is a partner of this project and allows us to meet the public.

This project has been granted 8Keuros from IRCICA.

8.2. National Initiatives

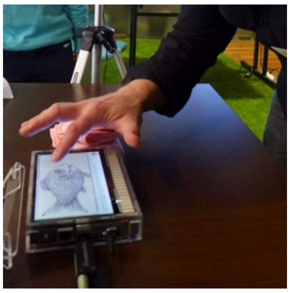
8.2.1. Touchit (13th FUI, May 2012-2015)

Participants: Michel Amberg, Frédéric Giraud, Betty Lemaire-Semail [correspondant].

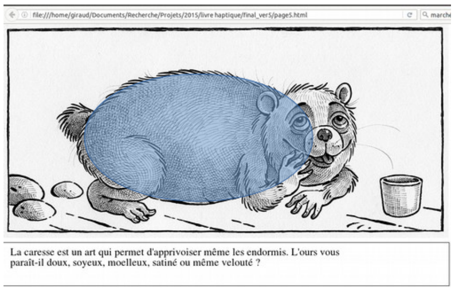
The purpose of this project is twofold. It aims at designing and implementing hardware solutions for tactile feedback based on programmable friction. It also aims at developing the knowledge and software tools required to use these new technologies for human-computer interaction. Grant for MINT is balanced on 272 keuro handled at University for L2EP, and 220 Keuros for Inria.

Partners: STMicroelectronics, CEA/LETI, Orange Labs, CNRS, EASii IC, MENAPIC and ALPHAUI.

Competitive clusters involved: Minalogic, Cap Digital and MAUD.



(a)



(b)

Figure 4. Demo session at "La nuit des Bibliothèques (Lille, October 2016), and a page of Stimtac; the ellipse highlights the tactile feedback on E-Vita.

8.2.2. Equipex IRDIVE (ANR project 2012-2020)

3 Meuros project, co-funded by ERDF for the developpement of a pluri-disciplinary project on ICT-based tools for understanding human perception of visual contents. Laurent Grisoni is member of the lead group of this project, and animates an axis devoted to art-sciences and technologies collaborations.

8.2.3. MAUVE CPER ("Contrat de Plan État-Région") 2016-2020 project

Funds: 4 Meuros (validated at national level, funded by Region), and 1 Meuro additional funding provided by ERDF.
Subject: ICT tools for mediation and access to knowledge.
Lead: University of Lille, University of Artois. Laurent Grisoni is co-lead of this project.

8.2.4. Projet FUI HID: lead Holusion (2016-2018)

Participants: Laurent Grisoni [correspondant], Samuel Degrande, Fabrice Aubert.

290 Keuros for MINT. Funding for two 18 months contracts and 24 months of post-doc.
Subject: rationalized process for industrial use of holographic displays.
MINT contribution: anamorphic software tools for holographics displays, and study of interactive aspects, including collaborative activities. This project has been prematurily stopped by french government.

8.2.5. InriaRT

Participants: Laurent Grisoni [correspondant], Samuel Degrande, Francesco de Comité.

Art/science Inria internal network gathering projects interested in collaborating with artists.
Inria teams involved: MuTANT (paris), Imagine (grenoble), Flowers, Potioc (Bordeaux), Hybrid, MimeTic (Rennes). This initiative shall take advantage of an agreement between Inria and french ministry of culture, signed early december 2016.

8.2.6. MATRICE (sept 2015-sept. 2017

Participant: Laurent Grisoni [correspondant].

This regional project, funded by ERDF, led by lille school of architecture, aims at understanding in which way 3D printing may be interesting for the building economy. partners: école d'architecture de Lille, Inria, ecole centrale de lille, télécome Lille 1, Ecole des mines de douai.

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

- INESC-ID: collaboration with Joaquim Jorge (Talaria paper, published at ISS'16)
- Collaboration with Mrad UofT (paper published, harvesting energy)

8.4. International Research Visitors

8.4.1. Visits of International Scientists

- Dr. Luke Dahl (University of Virginia) for the BOEUF project
- Masaya Takasaki and Masayuki Hara (University of Saitama, Japan) 22nd of january
- Masaya Takasaki has also been visiting Professor at University lille1 (April, 18th - April 30th)

8.4.1.1. Internships

visiting PhD student from University of Chile: Orlando Errazo (nov 2015-jan 2016). One publication currently on submission.

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. Member of the Conference Program Committees

Laurent Grisoni : PC for VISIGRAPP (IEEE InfoViz Art Track), Computer Graphics International (CGI, computer graphics), MOCO (international workshop on gesture), ISEA (art-science), GRAPP (computer graphics)

9.1.1.2. Reviewer

Florent Berthaut: Reviewer for ACM CHI Conference and NIME conference
Laurent Grisoni: Eurohaptics, ACM UIST, ACM CHI

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

Frédéric Giraud is Associate Editor of IEEE Transactions on Haptics

9.1.2.2. Reviewer - Reviewing Activities

Florent Berthaut: Reviewer for IEEE Multimedia
Laurent Grisoni: Computer & Graphics,

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

- Licence: Florent Berthaut, Spreadsheets et VBA programming, LEA L3 TCI (22h), Web programming, LEA L3, Université Lille 3, France
- Licence: Frédéric Giraud, Physique pour le génie électrique (30h), Université Lille 1
- Master: Christophe Giraud-Audine: Control of electrical machines (30h), Power Electronics (30h), signal processing (30h), niveau M1, ENSAM, France
- Master: Florent Berthaut, Web programming, LEA M1 TSM (16,5h) and LEA M1 RICI (48h), Database, LEA M1 ANI (22h), Université Lille 3, France
- Master : Frédéric Giraud, Control of electrical machines (30h), Power Electronics (40h) niveau M1, Université Lille1, France
- Master : Laurent Grisoni, NIHM : nouvelles Interactions Homme-Machine, (6h), niveau M2, Université Lille 1, France
- Master: Laurent Grisoni, représentation et compression de données (24h), introduction à la programmation (38h), cryptographie (8h), Ecole Polytech'lille (dept IMA)
- Master: Laurent Grisoni: gestion de projet en Co-design interdisciplinaire, Master Sciences et Culture du Visuel, Université de Lille Sciences Humaines et Sociales, Master Sciences et Cultures du Visuel (12h)
- Master : Laurent Grisoni, IHM et Interface à Gestes, (24h), niveau M2 (IMA5), Polytech Lille, France

9.2.2. Supervision

- PhD : Ehsan Enferad, Modélisation et Commande d'une Interface Tactile à Stimulation Hybride par Modulation de Friction et Retournement Temporel, nov. 2015, F. Giraud, C. Giraud-Audine
- PhD in progress: Cagan Arslan, Fusion de données pour l'interaction homme-machine, oct. 2015, L. Grisoni/J. Martinet
- PhD in progress: Farzan Kalantari, Interaction sur dispositif à retour tactile et kinesthésique, oct. 2014, L. Grisoni, F. Giraud
- PhD Sofiane Ghenna, Contrôle multimodal d'acionneurs piézo-électriques pour applications tactiles, F.Giraud, C.Giraud-Audine, University Lille1, 30th November 2016
- PhD Thomas Sednaoui, Tactile feedback integration on mobile communication devices, B Semail, F Casset, University Lille1, 14th dec. 2016
- PhD : Eric Vezzoli, Tactile feedback devices: friction control and texture generation, University Lille1, 22 Sept 2016, B. Lemaire-Semail, F. Giraud
- PhD in progress: Hanae Rateau, l'interaction esquissée, oct. 2012, L. Grisoni
- PhD in progress: Vincent Gouezou: L'architecte et ses outils, au travers de l'histoire et dans sa relation actuelle au numérique, oct. 2014, L. Grisoni 25% (with F. Vermandel, architect, Lille school of architecture)
- PhD in progress: Charlotte Planckeel, Le sens de la lacunae en archéologie de l'Âge du bronze, archéologie et outils numériques, L. Grisoni (25%, with A. Lehoerff, Lille 3, archeologist)
- PhD in progress: Olivier Capra, Interaction de présentation, oct. 2016, L. Grisoni, F. Berthaut

9.3. Popularization

- Futur en Seine (Paris, 9-10 june 2016; http://www.futur-en-seine.paris/). Presentation of E-Vita.
- Nuit des bibliothèques (Lille, 15th october 2016, http://www.bm-lille.fr/nuit-des-bibliotheques.aspx). Presentation of the haptic book to the readers of the public library.

- Forum du CNRS (Lille, 19th November 2016, http://leforum.cnrs.fr/). Presentation of 2 demos.
- Nuit des partenaires (Lille, 29th november 2016, https://www.univ-lille.fr/soireepartenaires/). Présentation of the Haptic book.
- Salon Humanités connectées, Toulouse, dec. 2016.
- Rencontres Inria Entreprises (25 novembre 2016), imaginarium.
- journées RV en haut de france, 24 novembre 2016, imaginarium.
- tables ronde journées régionales de l'innovation, amiens, 20 novembre 2016.
- BOEUF project workshops. Introduction to PureData and digital orchestras @ La condition publique (Roubaix) and l'Imaginarium (Tourcoing) http://www.cristal.univ-lille.fr/oeuvres-et-recherches/2016/11/28/workshop-les-orchestres-numeriques-espace-recherche-scv/

10. Bibliography

Major publications by the team in recent years

[1] F. BERTHAUT, M. HACHET.*Spatial Interfaces and Interactive 3D Environments for Immersive Musical Performances*, in "IEEE Computer Graphics and Applications", September 2016, vol. 36, n^{o} 5, p. 82 - 87 [*DOI :* 10.1109/MCG.2016.96], https://hal.inria.fr/hal-01374911.

[2] F. BERTHAUT, A. JONES.*ControllAR: Appropriation of Visual Feedback on Control Surfaces*, in "Interactive Spaces and Surfaces", Niagara Falls, Canada, Proceedings of ACM Interactive Spaces and Surfaces, ACM, November 2016 [*DOI :* 10.1145/2992154.2992170], https://hal.archives-ouvertes.fr/hal-01356239.

[3] M. BIET, F. GIRAUD, B. LEMAIRE-SEMAIL.*Squeeze film effect for the design of an ultrasonic tactile plate*, in "IEEE Transactions on Ultrasonic, Ferroelectric and Frequency Control", December 2007, vol. 54, n^{o} 12, p. 2678-2688, http://dx.doi.org/10.1109/TUFFC.2007.596.

[4] F. DE COMITE, L. GRISONI.*Numerical Anamorphosis: an Artistic Exploration*, in "SIGGRAPH ASIA 2015", Kobe, Japan, November 2015, https://hal.archives-ouvertes.fr/hal-01258727.

[5] H. RATEAU, L. GRISONI, B. DE ARAUJO.*Mimetic Interaction Spaces : Controlling Distant Displays in Pervasive Environments*, in "Intelligent User Interfaces", Haifa, Israel, February 2014 [*DOI :* 10.1145/2557500.2557545], https://hal.inria.fr/hal-01021337.

[6] H. RATEAU, Y. REKIK, L. GRISONI, J. JORGE.*Talaria: Continuous Drag & Drop on a Wall Display*, in "ISS'16", Niagara Falls, Canada, November 2016 [*DOI :* 10.1145/2992154.2992164], https://hal.inria.fr/hal-01381277.

[7] Y. REKIK, L. GRISONI, N. ROUSSEL.*Towards Many Gestures to One Command: A User Study for Tabletops*, in "INTERACT - 14th IFIP TC13 Conference on Human-Computer Interaction", Cape Town, South Africa, Springer, September 2013, https://hal.inria.fr/hal-00831877.

[8] Y. REKIK, R.-D. VATAVU, L. GRISONI.*Match-Up & Conquer: A Two-Step Technique for Recognizing Unconstrained Bimanual and Multi-Finger Touch Input*, in "AVI 2014, the 12th International Working Conference on Advanced Visual Interfaces", Como, Italy, May 2014, p. 201-208 [*DOI :* 10.1145/2598153.2598167], https://hal.inria.fr/hal-00991940.

[9] E. VEZZOLI, T. SEDNAOUI, M. AMBERG, F. GIRAUD, B. LEMAIRE-SEMAIL.*Texture Rendering Strategies with a High Fidelity - Capacitive Visual-Haptic Friction Control Device*, in "Haptics: Perception, Devices, Control, and Applications", London, United Kingdom, July 2016 [*DOI :* 10.1007/978-3-319-42321-0_23], https://hal.inria.fr/hal-01341981.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] W. BEN MESSAOUD.*Design and Control of a Tactile Stimulator for Real Texture Simulation: Application to Textile Fabrics*, Université Lille 1 : Sciences et Technologies, June 2016, https://tel.archives-ouvertes.fr/tel-01360590.

Articles in International Peer-Reviewed Journal

[11] W. BEN MESSAOUD, F. GIRAUD, B. LEMAIRE-SEMAIL, M. AMBERG, M.-A. BUENO.*Amplitude Control of an Ultrasonic Vibration for a Tactile Stimulator*, in "IEEE/ASME Transactions on Mechatronics", February 2016, vol. 21, n^{o} 3, p. 1692 - 1701 [*DOI :* 10.1109/TMECH.2016.2535300], https://hal.inria.fr/hal-01396831.

[12] F. BERTHAUT, L. DAHL.*BOEUF: A Unified Framework for Modeling and Designing Digital Orchestras*, in "Lecture notes in computer science", September 2016, vol. 9617, p. 153 - 166 [*DOI :* 10.1007/978-3-319-46282-0_10], https://hal.archives-ouvertes.fr/hal-01380373.

[13] F. BERTHAUT, M. HACHET.*Spatial Interfaces and Interactive 3D Environments for Immersive Musical Performances*, in "IEEE Computer Graphics and Applications", September 2016, vol. 36, n^{o} 5, p. 82 - 87 [*DOI :* 10.1109/MCG.2016.96], https://hal.inria.fr/hal-01374911.

[14] S. GHENNA, E. VEZZOLI, C. GIRAUD-AUDINE, F. GIRAUD, M. AMBERG, B. LEMAIRE-SEMAIL.*Enhancing Variable Friction Tactile Display using an ultrasonic travelling wave*, in "IEEE Transactions on Haptics (ToH)", September 2016, vol. Volume: PP, Issue: 99, https://hal.inria.fr/hal-01425156.

[15] V. KULKARNI, F. GIRAUD, C. GIRAUD-AUDINE, M. AMBERG, R. BEN MRAD, S. E. PRASAD.*Integration of a torsion-based shear-mode energy harvester and energy management electronics for a sensor module*, in "Journal of Intelligent Material Systems and Structures", November 2016 [*DOI :* 10.1177/1045389X16672563], https://hal.inria.fr/hal-01396823.

International Conferences with Proceedings

[16] F. BERTHAUT, A. JONES.*ControllAR: Appropriation of Visual Feedback on Control Surfaces*, in "Interactive Spaces and Surfaces", Niagara Falls, Canada, Proceedings of ACM Interactive Spaces and Surfaces, ACM, November 2016 [*DOI :* 10.1145/2992154.2992170], https://hal.archives-ouvertes.fr/hal-01356239.

[17] C. DURIEZ, E. COEVOET, F. LARGILLIERE, T. M. BIEZE, Z. ZHANG, M. SANZ-LOPEZ, B. CARREZ, D. MARCHAL, O. GOURY, J. DEQUIDT.*Framework for online simulation of soft robots with optimization-based inverse model*, in "SIMPAR: IEEE International Conference on Simulation, Modeling, and Programming for Autonomous Robots", San Francisco, United States, Proceedings of SIMPAR 2016 conference, December 2016, https://hal.inria.fr/hal-01425349.

[18] F. LARGILLIÈRE, E. COEVOET, M. SANZ-LOPEZ, L. GRISONI, C. DURIEZ.*Stiffness rendering on soft tangible devices controlled through inverse FEM simulation*, in "International Conference on Intelligent Robots and Systems - IROS 2016", Daejeon, South Korea, October 2016, https://hal.inria.fr/hal-01386787.

[19] H. RATEAU, Y. REKIK, L. GRISONI, J. JORGE.*Talaria: Continuous Drag & Drop on a Wall Display*, in "ISS'16", Niagara Falls, Canada, November 2016 [*DOI :* 10.1145/2992154.2992164], https://hal.inria.fr/hal-01381277.

[20] T. SEDNAOUI, E. VEZZOLI, D. GUEORGUIEV, C. CHAPPAZ, B. LEMAIRE-SEMAIL.*Psychophysical Power Optimization of Friction Modulation for Tactile Interfaces*, in "Haptics: Perception, Devices, Control, and Applications", London, United Kingdom, July 2016 [*DOI :* 10.1007/978-3-319-42324-1_35], https://hal.inria.fr/hal-01342010.

[21] Z. VIDRIH, E. VEZZOLI.*Electrovibration Signal Design : A Simulative Approach*, in "Haptics: Perception, Devices, Control, and Applications", London, United Kingdom, July 2016, https://hal.inria.fr/hal-01341974.

Conferences without Proceedings

[22] F. KALANTARI, L. GRISONI, F. GIRAUD, Y. REKIK.*Finding the Minimum Perceivable Size of a Tactile Element on an Ultrasonic Based Haptic Tablet*, in "ISS '16 Extended Abstract, 11th ACM International Conference on Interactive Surfaces and Spaces", Niagara Falls, ON, Canada, November 2016, 6 [*DOI :* 10.1145/2992154.2996785], https://hal.archives-ouvertes.fr/hal-01381314.

[23] A. MOUNGOU, E. VEZZOLI, C. LOMBART, B. LEMAIRE-SEMAIL, J.-L. THONNARD, A. MOURAUX.*A novel method using EEG to characterize the cortical processes involved in active and passive touch*, in "IEEE - Haptic Symposium 2016", Philadelphia, United States, April 2016 [*DOI :* 10.1109/HAPTICS.2016.7463178], https://hal.inria.fr/hal-01318133.

[24] E. VEZZOLI, T. SEDNAOUI, M. AMBERG, F. GIRAUD, B. LEMAIRE-SEMAIL.*Texture Rendering Strategies with a High Fidelity - Capacitive Visual-Haptic Friction Control Device*, in "Haptics: Perception, Devices, Control, and Applications", London, United Kingdom, July 2016 [*DOI :* 10.1007/978-3-319-42321-0_23], https://hal.inria.fr/hal-01341981.

Scientific Books (or Scientific Book chapters)

[25] S. GHENNA, C. GIRAUD-AUDINE, F. GIRAUD, M. AMBERG, B. SEMAIL.*Modal Superimposition for Multifingers Variable Friction Tactile Device*, in "Haptics: Perception, Devices, Control, and Applications", July 2016, vol. 9774, p. 521-530 [*DOI :* 10.1007/978-3-319-42321-0_49], https://hal.inria.fr/hal-01360917.

References in notes

[26] M. BEAUDOUIN-LAFON.*Designing interaction, not interfaces*, in "Proceedings of AVI'04", ACM, 2004, p. 15-22, http://doi.acm.org/10.1145/989863.989865.

[27] M. BEAUDOUIN-LAFON, W. E. MACKAY.*Prototyping tools and techniques*, in "The Human Computer Interaction handbook: fundamentals, evolving technologies and emerging applications", A. SEARS, J. A. JACKO (editors), CRC Press, 2007, p. 1017-1039.

[28] G. CASIEZ, D. VOGEL, R. BALAKRISHNAN, A. COCKBURN.*The Impact of Control-Display Gain on User Performance in Pointing Tasks*, in "Human-Computer Interaction", 2008, vol. 23, n^{o} 3, p. 215–250.

[29] W. E. MACKAY.*Users and customizable software: a co-adaptive phenomenon*, Massachusetts Institute of Technology, May 1990.

[30] M. J. MULLER.*Participatory design: the third space in HCI*, in "The Human Computer Interaction handbook: fundamentals, evolving technologies and emerging applications", A. SEARS, J. A. JACKO (editors), CRC Press, 2007, p. 1061-1081.

[31] D. H. RUBINE.*The automatic recognition of gestures*, Carnegie Mellon University, December 1991, http://www.cs.cmu.edu/~music/papers/dean_rubine_thesis.ps.

[32] W. R. SHADISH, T. D. COOK, D. T. CAMPBELL.*Experimental and quasi-experimental designs for generalized causal inference*, Wadsworth Publishing, 2002.

[33] A. D. WILSON, A. F. BOBICK.*Parametric Hidden Markov Models for gesture recognition*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", September 1999, vol. 21, n^{o} 9, p. 884–900.

[34] J. O. WOBBROCK, A. D. WILSON, Y. LI.*Gestures without libraries, toolkits or training: a $1 recognizer for user interface prototypes*, in "Proceedings of UIST'07", ACM, 2007, p. 159–168, http://doi.acm.org/10.1145/1294211.1294238.

[35] G. A. TEN HOLT, M. J. T. REINDERS, E. A. HENDRIKS.*Multi-dimensional dynamic time warping for gesture recognition*, in "Proceedings of the Thirteenth annual conference of the Advanced School for Computing and Imaging", 2007, http://mediamatica.ewi.tudelft.nl/content/multi-dimensional-dynamic-time-warping-gesture-recognition.

Team Mjolnir

Computing tools to empower users

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Lille - Nord Europe

THEME
Interaction and visualization

Table of contents

Team Mjolnir

Creation of the Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

5.1. - Human-Computer Interaction
5.1.1. - Engineering of interactive systems
5.1.2. - Evaluation of interactive systems
5.1.3. - Haptic interfaces
5.1.4. - Brain-computer interfaces, physiological computing
5.1.5. - Body-based interfaces
5.1.8. - 3D User Interfaces
5.2. - Data visualization
5.5.4. - Animation
5.7.2. - Music

Other Research Topics and Application Domains:

1.1.5. - Genetics
2.2.6. - Neurodegenerative diseases
2.8. - Sports, performance, motor skills
6.1.1. - Software engineering
6.5. - Information systems
9.2.1. - Music, sound
9.2.2. - Cinema, Television
9.2.3. - Video games
9.4.1. - Computer science
9.5.5. - Sociology
9.5.10. - Digital humanities

1. Members

Research Scientists

Nicolas Roussel [Team leader, Inria, Senior Researcher, HDR]
Stéphane Huot [Inria, Senior Researcher, HDR]
Fanny Chevalier [Inria, Researcher]
Sylvain Malacria [Inria, Researcher]
Mathieu Nancel [Inria, Researcher, since November]
Marcelo Wanderley [Inria, International Chair]

Faculty Members

Géry Casiez [Université Lille 1, Professor, HDR]
Thomas Pietrzak [Université Lille 1, Assistant Professor]
Edward Lank [Université Lille 1 & University of Waterloo, Assistant Professor, since September]

Technical Staff

Izzatbek Mukhanov [Inria, until September]
Sébastien Poulmane [Inria]

Julien Decaudin [Inria, since November]

PhD Students

Alix Goguey [Inria, until October]
Justin Mathew [Inria]
Amira Chalbi-Neffati [Inria]
Thibault Raffaillac [Inria]
Nicole Pong [Inria, since October]

Post-Doctoral Fellow

Christian Frisson [Inria, since June]

Visiting Scientists

Filipe Calegario [McGill University, PhD student, in January]
Nicholas Fellion [Carleton University, Master's student, from January to April]
Daniel Wigdor [University of Toronto, Associate Professor, in April]
Hrim Mehta [Ontario Institute of Technology, PhD student, from May to August]
Aakar Gupta [University of Toronto, PhD student, from June to September]
Anastasia Kuzminykh [University of Waterloo, PhD Student, from October to December]
Baptiste Caramiaux [McGill University & IRCAM, PostDoc, in December]

Administrative Assistant

Karine Lewandowski [Inria]

2. Overall Objectives

2.1. Introduction

Human-Computer Interaction (HCI) is a constantly moving field. Changes in computing technologies extend their possible uses and modify the conditions of existing ones. People also adapt to new technologies and adapt them to their own needs. Different problems and opportunities thus regularly appear. Over the recent years, though, we believe incremental news have unfortunately eclipsed fundamental HCI topics on which a lot of work remains to be done. In what follows, we summarize the essential elements of our vision and the associated long-term goals.

2.2. Computers as tools

In the early 1960s, at a time where computers were scarce, expensive, bulky and formal-scheduled machines used for automatic computations, Engelbart saw their potential as personal interactive resources. He saw them as *tools*, as things we would purposefuly use to carry out particular tasks [48]. Others at the same time had a different vision. They saw computers as *partners*, intelligent entities to whom we would delegate tasks. These two visions constitute the roots of today's predominant human-computer interaction paradigms, *use* and *delegation*. Our focus is on computer users and our work should ultimately benefit them. Our interest is not in solving the difficult problems related to machine understanding. It is not in what machines understand, but in what people can do with them. Instead of intelligent systems, we aim for systems supporting intelligent use and empowering people. We do not reject the delegation paradigm but clearly favor the one of tool use.

2.3. Tools supporting transparent use

Technology is most empowering when it is transparent. But the transparent tool is not the one you cannot see, it is the one invisible in effect, the one that does not get into your way but lets you focus on the task. Heidegger used the term *zuhanden* (*ready-to-hand*) to characterize this unobtruded relation to things [50]. Transparency of interaction is not best achieved with tools mimicking human capabilities, but with those taking full advantage of them and fitted to the context and task. Our actions towards the digital world need to be digitized, and the digital world must provide us with proper feedbacks in return. Input and output technologies pose inevitable constraints while the digital world calls for more and more sophisticated perception-action couplings for increasingly complex tasks. We want to study the means currently available for perception and action in the digital world. We understand the important role of the body on the human side, and the importance of hardware elements on the computer side. Our work thus follows a systems approach encompassing these elements and all the software layers above, from device drivers to applications.

2.4. But tools also designed for analytic use

Engelbart believed in the coevolution of humans and their tools. He was not just interested in designing a personal computer but also in changing people, to radically improve the way we manage complexity. The human side of this coevolutionary process has been largely ignored by the computing industry which has focused on the development of walk-up-and-use interfaces for novice users. As a result of this focus on initial performance, we are trapped in a "beginner mode" of interaction with a low performance ceiling [5]. People find it acceptable to spend considerable amounts of time learning and practising all sorts of skills. We want to tap into these resources to develop digital skills. We must accept that new powerful tools might not support immediate transparent use and thus require attention. Heidegger used the term *vorhanden* (present-at-hand) to characterize the analytic relation to things that not only occurs when learning about them, but also when handling breakdowns, when they change or need to be adapted, or when teaching others how to use them. Analytic use is unavoidable and its interplay with transparent use is essential to tool accommodation and appropriation [47]. We want to study this interplay.

3. Research Program

3.1. Introduction

Our research program is organized around three main themes: leveraging human control skills, leveraging human perceptual skills, and leveraging human learning skills.

3.2. Leveraging human control skills

Our group has developed a unique and recognized expertise in *transfer functions*, i.e. the algorithmic transformations of raw user input for system use. Transfer functions define how user actions are taken into account by the system. They can make a task easier or impossible and thus largely condition user performance, no matter the criteria (speed, accuracy, comfort, fatigue, etc). Ideally, the transfer function should be chosen or tuned to match the interaction context. Yet the question of how to design a function to maximize one or more criteria in a given context remains an open one, and on-demand adaptation is difficult because functions are usually implemented at the lowest possible level to avoid latency problems. Latency management and transfer function design are two problems that require cross examination to improve human performance with interactive systems. Both also contribute to the senses of *initiation* and *control*, two crucial component of the sense of *agency* [51]. Our ultimate goal on these topics is to adapt the transfer function to the user and task in order to support stable and appropriate control. To achieve this, we investigate combinations of low-level (embedded) and high-level (application) ways to take user capabilities and task characteristics into account and reduce or compensate for latency in different contexts, e.g. using a mouse or a touchpad, a touch-screen, an optical finger navigation device or a brain-computer interface.

3.3. Leveraging human perceptual skills

Our work under this theme concerns the physicality of human-computer interaction, with a focus on haptic perception and related technologies, and the perception of animated displays.

Vibrators have long been used to provide basic kinesthetic feedback. Other piezoceramic and electro-active polymer technologies make it possible to support programmable friction or emboss a surface, and thin, organic technologies should soon provide transparent and conformable, flexible or stretchable substrates. We want to study the use of these different technologies for static and dynamic haptic feedback from both an engineering and an HCI perspective. We want to develop the tools and knowledge required to facilitate and inform the design of future haptic interactions taking best advantage of the different technologies.

Animations are increasingly common in graphical interfaces. Beyond their compelling nature, they are powerful tools that can be used to depict dynamic data, to help understand time-varying behaviors, to communicate a particular message or to capture attention. Yet despite their popularity, they are still largely under-comprehended as cognitive aids. While best practices provide useful directions, very little empirical research examine different types of animation, and their actual benefits and limitations remain to be determined. We want to increase current knowledge and develop the tools required to best take advantage of them.

3.4. Leveraging human learning skills

By looking at ways to leverage human control and perceptual skills, the research yet proposed mainly aims at improving perception-action coupling to better support transparent use. This third research theme addresses the different and orthogonal topic of skill acquisition and improvement. We want to move away from the usual binary distinction between “novices” and “experts” and explore means to promote and assist digital skill development in a more progressive fashion. We are interested in means to support the analytic use of computing tools. We want to help people become aware of the particular ways they use their tools, the other ways that exist for the things they do, and the other things they might do. We want to help them increase their performance by adjusting their current ways of doing, by providing new and more efficient ways, and by facilitating transitions from one way to another. We are also interested in means to foster reflection among users and facilitate the dissemination of best practices.

4. Application Domains

4.1. Application Domains

Mjolnir works on fundamental aspects of Human-Computer Interaction that can be applied to diverse application domains. Our 2016 research concerned desktop and touch-based interfaces with notable applications to social network analysis, genetics research, 3D environments, as well as 3D films and Virtual Reality stories.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Personnel

Mathieu Nancel joined us as an Inria researcher in November.

Marcelo Wanderley joined us in February as part of the Inria International Chair program and will spend 20% of his time with us until 2020.

Ed Lank, Associate Professor at the University of Waterloo, joined us in September for a long-term visit (10+ months) funded by Région Hauts-de-France and Université Lille 1.

In partnership with Campus France and Inria, Mitacs' Globalink Research Award program sponsored the visits of three canadian students in our group: Nicholas Fellion (Carleton University), Hrim Mehta (Ontario Institute of Technology) and Aakar Gupta (University of Toronto).

5.1.2. Publications & dissemination

Mjolnir presented seven papers and one "late-breaking work" at the ACM CHI 2016 conference in May, the most prestigious conference in our field.

The *Animated transitions* web site launched in March illustrates previous works by Fanny Chevalier and others on this topic (Histomages, Diffamation and Gliimpse).

5.1.3. Awards

"Honorable mention" (top 5% of the 2300+ submissions) from the ACM CHI 2016 conference to the following three papers:

- "Egocentric analysis of dynamic networks with EgoLines", from J. Zhao, M. Glueck, F. Chevalier, Y. Wu & A. Khan
- "Modeling and understanding human routine behavior", from N. Banovic, T. Buzali, F. Chevalier, J. Mankoff & A. Dey
- "Direct manipulation in tactile displays", from A. Gupta, T. Pietrzak, N. Roussel & R. Balakrishnan

"Springer award for best doctoral contribution" to Amira Chalbi-Neffati at the IHM 2016 conference.
BEST PAPERS AWARDS :
[40] **CHI '16 - Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. J. ZHAO, M. GLUECK, F. CHEVALIER, Y. WU, A. KHAN.
[23] **ACM CHI Conference on Human Factors in Computing Systems 2016**. N. BANOVIC, T. BUZALI, F. CHEVALIER, J. MANKOFF, A. K. DEY.
[29] **CHI 2016 - Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems**. A. GUPTA, T. PIETRZAK, N. ROUSSEL, R. BALAKRISHNAN.

6. New Software and Platforms

6.1. Introduction

Each software listed below is characterized according to the *criteria for software self-assessment* proposed by Inria's Evaluation Committee. Note that the only software mentioned here are those that were created or significantly modified during the year.

6.2. Libpointing

Participants: Géry Casiez [correspondent], Nicolas Roussel, Izzatbek Mukhanov, Sébastien Poulmane.

Libpointing is a software toolkit that provides direct access to HID pointing devices and supports the design and evaluation of pointing transfer functions [3]. The toolkit provides resolution and frequency information for the available pointing and display devices and makes it easy to choose between them at run-time through the use of URIs. It allows to bypass the system's transfer functions to receive raw asynchronous events from one or more pointing devices. It replicates as faithfully as possible the transfer functions used by Microsoft Windows, Apple OS X and Xorg (the X.Org Foundation server). Running on these three platforms, it makes it possible to compare the replicated functions to the genuine ones as well as custom ones. The toolkit is written in C++ with Python, Java and Node.js bindings available (about 49,000 lines of code in total). It is publicly available under the GPLv2 license.

The library has been thoroughly improved in 2016. Notable changes concern the migration of code on GitHub, the set up of continuous integration and the automated release of buildings for Windows, Linux and MacOS. libpointing can be easily installed using apt-get command on Linux and and Homebrew and Macport package installers on MacOS. New features like the estimation of the input frequency have been added and different demos have been developed. Code has been re-factored and various bugs fixed.

Web site: http://libpointing.org/

Software characterization: [A-3] [SO-3] [SM-2] [EM-2↑] [SDL-5]

6.3. Liblag

Participants: Géry Casiez [correspondent], Nicolas Roussel, Sébastien Poulmane.

Liblag is a software toolkit designed to support the comparison of latency compensation techniques. The toolkit notably includes a playground application that allows to compare different trajectory prediction algorithms on desktop (OS X, Ubuntu and Windows) and mobile (iOS and Android) systems. The source code for this toolkit (about 8,500 lines of code) is only available to Turbotouch partners for now.

Sébastien Poulmane was recruited in May as an engineer on the TurboTouch project. He has been contributing to refactor the code, integrate new input devices and new prediction techniques and also develop associated demos and experiments.

Software characterization: [A-1] [SO-4] [SM-1] [EM-2] [SDL-1]

6.4. Mouse-based lagmeter

Participants: Géry Casiez [correspondent], Stéphane Huot, Nicolas Roussel.

As part of the work reported in [46], we implemented our mouse-based method for measuring end-to-end latency using Java/Swing, C++/GLUT, C++/Qt and JavaScript/HTML5. We also wrote Python scripts to parse the logs generated by these implementations in order to compare them. This software (about 2,500 lines of code) was made available in 2016 on a public git repository. The online interactive demo has been improved to collect anonymous latency measurement data from users and integrate libpointing in order to get information about the input and output devices connected. A native Android version has also been developed.

Web site: http://mjolnir.lille.inria.fr/turbotouch/lagmeter/

Software characterization: [A-1] [SO-4] [SM-1] [EM-2] [SDL-1]

6.5. TAN

Participants: Fanny Chevalier [correspondent], Nicolas Roussel, Stéphane Huot.

TAN stands for *Transitions animées*, i.e. *Animated transitions*. This web site illustrates some of our past research on this topic. It combines videos and live demonstrations of Histomages, an image editing tool that associates pixel and color space; Diffamation, an animation tool to follow and understand the modifications made to a document over time; and Gliimpse, a markup language editor (e.g. HTML, LaTeX, Wiki) to instantly switch from source code to the document it produces and vice versa. The source code for the three demonstrators (about 87,000 lines of Java and JavaScript) is not distributed for the moment.

Web site: http://tan.lille.inria.fr/

Software characterization: [A-4] [SO-2] [SM-3] [EM-2] [SDL-4]

6.6. InspectorWidget

Participants: Christian Frisson [correspondent], Sylvain Malacria.

InspectorWidget [26] is an HTML5/nodejs/C++ software suite that can be used by an experimenter to track and analyze users' behaviors in closed interactive software. The suite has a *recording* module that records users' display and captures low-level events while she carries her task, and an *annotation* module that combines OCR and low-level inputs analysis so the experimenter post-annotate users' activity afterwards. InspectorWidget is cross-platform, open-source and publicly available under the GPLv3 license. New features, notably recording and exploiting accessibility API, are currently under development in order to be tested and added to the software suite.

Web site: https://github.com/InspectorWidget/InspectorWidget/

Software characterization: [A-2↑] [SO-3] [SM-3↑] [EM-3] [SDL-4]

7. New Results

7.1. Introduction

The following sections summarize our main results of the year. For a complete list, see the list of publications at the end of this report.

7.2. Understanding and modeling users

Participants: Géry Casiez, Christian Frisson, Alix Goguey, Stéphane Huot, Sylvain Malacria, Mathieu Nancel, Thibault Raffaillac, Nicolas Roussel.

7.2.1. Touch interaction with finger identification: which finger(s) for what?

The development of robust methods to identify which finger is causing each touch point, called "finger identification," will open up a new input space where interaction designers can associate system actions to different fingers [11]. However, relatively little is known about the performance of specific fingers as single touch points or when used together in a "chord". We presented empirical results for accuracy, throughput, and subjective preference gathered in five experiments with 48 participants exploring all 10 fingers and 7 two-finger chords. Based on these results, we developed design guidelines for reasonable target sizes for specific fingers and two-finger chords, and a relative ranking of the suitability of fingers and two-finger chords for common multi-touch tasks. Our work contributes new knowledge regarding specific finger and chord performance and can inform the design of future interaction techniques and interfaces utilizing finger identification [28].

7.2.2. Training and use of brain-computer interfaces

Brain-Computer Interfaces (BCIs) are much less reliable than other input devices, with error rates ranging from 5% up to 60%. To assess the subjective frustration, motivation, and fatigue of users when confronted to different levels of error rate, we conducted an BCI experiment in which it was artificially controlled. Our results show that a prolonged use of BCI significantly increases the perceived fatigue, and induces a drop in motivation [38]. We also found that user frustration increases with the error rate of the system but this increase does not seem critical for small differences of error rate. For future BCIs, we thus advise to favor user comfort over accuracy when the potential gain of accuracy remains small.

We have also investigated if the stimulation used for training an SSVEP-based BCI have to be similar to the one used *in fine* for interaction. We recorded 6-channels EEG data from 12 subjects in various conditions of distance between targets, and of difference in color between targets. Our analysis revealed that the stimulation configuration used for training which leads to the best classification accuracy is not always the one which is closest to the end use configuration [15]. We found that the distance between targets during training is of little influence if the end use targets are close to each other, but that training at far distance can lead to a better accuracy for far distance end use. Additionally, an interaction effect is observed between training and testing color: while training with monochrome targets leads to good performance only when the test context involves monochrome targets as well, a classifier trained on colored targets can be efficient both for colored and monochrome targets. In a nutshell, in the context of SSVEP-based BCI, training using distant targets of different colors seems to lead to the best and more robust performance in all end use contexts.

7.2.3. Evaluation metrics for touch latency compensation

Touch systems have a delay between user input and corresponding visual feedback, called input “latency” (or “lag”). Visual latency is more noticeable during continuous input actions like dragging, so methods to display feedback based on the most likely path for the next few input points have been described in research papers and patents. Designing these “next-point prediction” methods is challenging, and there have been no standard metrics to compare different approaches. We introduced metrics to quantify the probability of 7 spatial error “side-effects” caused by next-point prediction methods [35]. Types of side-effects were derived using a thematic analysis of comments gathered in a 12 participants study covering drawing, dragging, and panning tasks using 5 state-of-the-art next-point predictors. Using experiment logs of actual and predicted input points, we developed quantitative metrics that correlate positively with the frequency of perceived side-effects. These metrics enable practitioners to compare next-point predictors using only input logs.

7.2.4. Application use in the real world

Interface designers, HCI researchers or usability experts often need to collect information regarding usage of interactive systems and applications in order to interpret quantitative and behavioral aspects from users – such as our study on the use of trackpads described before – or to provide user interface guidelines. Unfortunately, most existing applications are closed to such probing methods: source code or scripting support are not always available to collect and analyze users’ behaviors in real world scenarios.

InspectorWidget [26] is an open-source cross-platform application we designed to track and analyze users’ behaviors in interactive software. The key benefits of this application are: 1) it works with closed applications that do not provide source code nor scripting capabilities; 2) it covers the whole pipeline of software analysis from logging input events to visual statistics through browsing and programmable annotation; 3) it allows post-recording logging; and 4) it does not require programming skills. To achieve this, InspectorWidget combines low-level event logging (e.g. mouse and keyboard events) and high-level screen capturing and interpretation features (e.g. interface widgets detection) through computer vision techniques.

7.2.5. Trackpad use in the real world

Trackpads (or *touchpads*) allow to control an on-screen cursor with finger movements on their surface. Recent models also support force sensing and multi-touch interactions, which make it possible to scroll a document by moving two fingers or to switch between virtual desktops with four fingers, for example. But despite their widespread use, little is known about how users interact with them, and which gestures they are most familiar with. To better understand this, we conducted a three-steps field study with Apple Macbook’s multi-touch trackpads.

The first step of our study consisted in collecting low-level interaction data such as contact points with the trackpad and the multi-touch gestures performed while interacting. We developed a dedicated interaction logging application that we deployed on the workstation of 11 users for a duration of 14 days, and collected a total of over 82 millions contact points and almost 220 000 gestures. We then investigated finger chords (i.e., fingers used) and hand usage when interacting with a trackpad. In that purpose, we designed a dedicated mirror stand that can be easily positioned in front of the laptop’s embedded web camera to divert its capturing field

(Figure 1, left). This mirror stand is combined with a background application taking photos when a multi-finger gesture is performed. We deployed this setup on the computer of 9 users for a duration of 14 days. Finally, we deployed a system preference collection application to gather the trackpad system preferences (such as transfer function and gestures associated) of 80 users. Our main findings are that touch contacts on the trackpad are performed on a limited sub-surface and are relatively slow (Figure 1, right); that the consistency of user finger chords varies depending on the frequency of a gesture and the number of fingers involved; and that users tend to rely on the default system preferences of the trackpad [34].

Figure 1. Left: Mirror positioned in front of a built-in camera to divert its field of view and analyze finger-chords usages; Right: frequency distribution of all touch events of our participants, overlaid on a trackpad.

7.3. Interactive visualization and animations

Participants: Amira Chalbi-Neffati, Fanny Chevalier, Nicolas Roussel.

7.3.1. *Social network analysis*

The egocentric analysis of dynamic networks focuses on discovering the temporal patterns of a subnetwork around a specific central actor, i.e. an ego-network. These types of analyses are useful in many application domains, such as social science and business intelligence, providing insights about how the central actor interacts with the outside world. *EgoLines* is an interactive visualization we designed to support the egocentric analysis of dynamic networks . Using a “subway map” metaphor, a user can trace an individual actor over the evolution of the ego-network (Figure 2). The design of EgoLines is grounded in a set of key analytical questions pertinent to egocentric analysis, derived from interviews with three domain experts and general network analysis tasks. The results of controlled experiments with end-users and domain experts show its effectiveness in egocentric analysis tasks. Egolines can be tested at http://fannychevalier.net/egolines.html

7.3.2. *Cross-sectional cohort phenotype*

Cross-sectional phenotype studies are used by genetics researchers to better understand how phenotypes vary across patients with genetic diseases, both within and between cohorts. Analyses within cohorts identify patterns between phenotypes and patients, e.g. co-occurrence, and isolate special cases, e.g. potential outliers). Comparing the variation of phenotypes between two cohorts can help distinguish how different factors affect disease manifestation, e.g. causal genes, age of onset.). *PhenoStacks* is a novel visual analytics tool we designed to support the exploration of phenotype variation within and between cross-sectional patient cohorts . By leveraging the semantic hierarchy of the Human Phenotype Ontology, phenotypes are presented in context, can be grouped and clustered, and are summarized via overviews to support the exploration of phenotype distributions (Figure 3). The HPO is rarely used for visualization and was only recently first employed in PhenoBlocks [49]. In this prior work, we used the HPO to visualize phenotypes in clinical diagnosis settings, supporting the pairwise comparison of patient phenotypes using explicit encoding. In

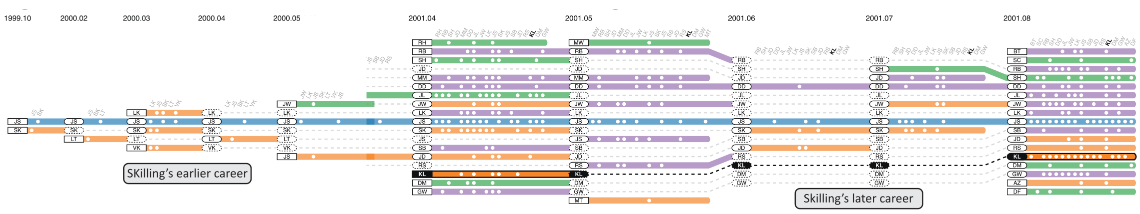


Figure 2. Egolines used to explore the dynamic network of email communications among employees at the Enron company.

this new work, we turn our focus to genetics researchers conducting cross-sectional cohort studies, where the distribution of phenotypes is compared across many patients. The design of PhenoStacks was motivated by formative interviews with genetics researchers. The results of a deployment evaluation with four expert genetics researchers suggest that PhenoStacks can help identify phenotype patterns, investigate data quality issues, and inform data collection design. PhenoStacks is available from http://phenostacks.org/

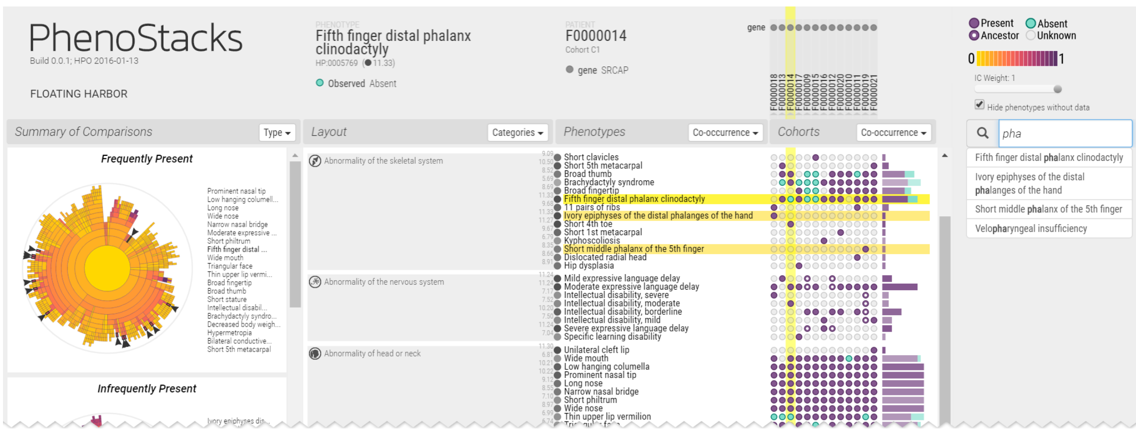


Figure 3. Exploration of phenotypic variation in cross-sectional cohorts of patients with a rare genetic disease using PhenoStacks.

7.3.3. Human routine behavior

Human routines are blueprints of behavior, which allow people to accomplish purposeful repetitive tasks at many levels, ranging from the structure of their day to how they drive through an intersection. People express their routines through actions that they perform in the particular situations that triggered those actions. An ability to model routines and understand the situations in which they are likely to occur could allow technology to help people improve their bad habits, inexpert behavior, and other suboptimal routines. However, existing routine models do not capture the causal relationships between situations and actions that describe routines. Byproducts of an existing activity prediction algorithm can be used to model those causal relationships in routines [23]. We applied this algorithm on two example datasets, and showed that the modeled routines are meaningful — that they are predictive of people's actions and that the modeled causal relationships provide insights about the routines that match findings from previous research. Our approach offers a generalizable solution to model and reason about routines. We show that the extracted routine patterns are at least as predictive of behaviors in the two behavior logs as the baseline we establish with existing algorithms.

To make the routine behavior models created using our approach accessible to participants and allow them to investigate the extracted routine patterns, we developed a simple visualization tool. To maintain a level of familiarity, we base our visual encoding of routine behavior elements on a traditional visual representation of an MDP as a graph (Figure 4). Our MDP graph contains nodes representing states (as circles) and actions (as squares), directed edges from state nodes to action nodes (indicating possible actions people can perform in those states), and directed edges from actions to states (indicating state transitions for any given state and action combination).

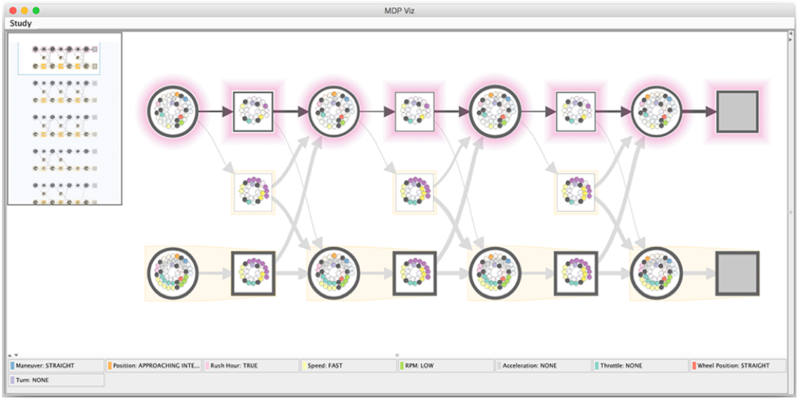


Figure 4. Our visual analytics tool showing the main routine and one likely variation of non-aggressive drivers extracted using our approach.

7.3.4. Meta-analysis of data based on user-authored annotations

User-authored annotations of data can support analysts in the activity of hypothesis generation and sense-making, where it is not only critical to document key observations, but also to communicate insights between analysts. *Annotation Graphs* are dynamic graph visualizations that enable meta-analysis of data based on user-authored annotations . The annotation graph topology encodes annotation semantics, which describe the content of and relations between data selections, comments, and tags. We present a mixed-initiative approach to graph layout that integrates an analyst's manual manipulations with an automatic method based on similarity inferred from the annotation semantics. Annotation graphs are implemented within a system, C8, that supports authoring annotations during exploratory analysis of a dataset (Figure 5). In this work, we develop and evaluate the system through an iterative user-centered design process with three experts, situated in the domain of analyzing HCI experiment data. The results suggest that annotation graphs are effective as a method of visually extending user-authored annotations to data meta-analysis for discovery and organization of ideas.

7.3.5. Fundamentals of animated transitions

Animations are increasingly used in interactive systems in order to enhance the usability and aesthetics of user interfaces. While animations are proven to be useful in many cases, we still find defective ones causing many problems, such as distracting users from their main task or making data exploration slower. The fact that such animations still exist proves that animations are not yet very well understood as a cognitive aid, and that we have not yet definitely decided what makes a well designed one. Our work on this topic aims at better understanding the different aspects of animations for user interfaces and exploring new methods and guidelines for designing them.

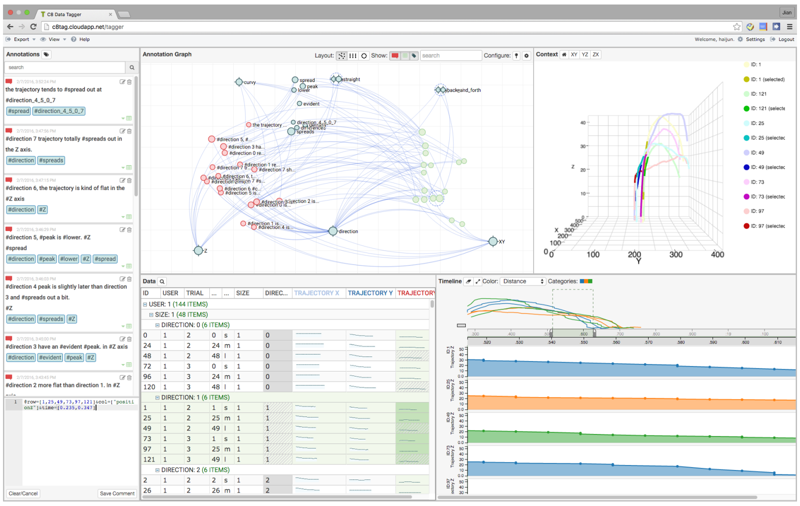


Figure 5. C8 used to analyze the results of an HCI user study that records participants pointing at a target on a tabletop display with different experimental conditions.

From bouncing icons that catch attention, to transitions helping with orientation, to tutorials, animations can serve numerous purposes. In , we revisit Baecker and Small's pioneering work *Animation at the Interface*, 25 years later. We review academic publications and commercial systems, and interviewed 20 professionals of various backgrounds. Our insights led to an expanded set of roles played by animation in interfaces today for keeping in context, teaching, improving user experience, data encoding and visual discourse. We illustrate each role with examples from practice and research, discussed evaluation methods and point to opportunities for future research. This expanded description of roles aims at inspiring the HCI research community to find novel uses of animation, guide them towards evaluation and spark further research.

We have also studied different aspects of animations for visual analysis tasks. We have worked on the design of a new model for animated transitions, explored certain aspects of visual grouping for these transitions, and studied the impact of their temporal structure on data interpretation. These works, while still in progress, have been presented at the IHM doctoral consortium [39].

7.4. Interaction techniques

Participants: Géry Casiez, Fanny Chevalier, Stéphane Huot, Sylvain Malacria, Justin Mathew, Thomas Pietrzak, Nicolas Roussel.

7.4.1. Interaction in 3D environments

In virtual environments, interacting directly with our hands and fingers greatly contributes to the sense of immersion, especially when force feedback is provided for simulating the touch of virtual objects. Yet, common haptic interfaces are unfit for multi-finger manipulation and only costly and cumbersome grounded exoskeletons do provide all the efforts expected from object manipulation. To make multi-finger haptic interaction more accessible, we propose to combine two affordable haptic interfaces into a bimanual setup named DesktopGlove [18]. With this approach, each hand is in charge of different components of object manipulation: one commands the global motion of a virtual hand while the other controls its fingers for grasping. In addition, each hand is subjected to forces that relate to its own degrees of freedom so that users perceive a variety of haptic effects through both of them. Our results show that (1) users are able to integrate the separated degrees of freedom of DesktopGlove to efficiently control a virtual hand in a posing task, (2) DesktopGlove shows overall better performance than a traditional data glove and is preferred by users, and (3) users considered the separated haptic feedback realistic and accurate for manipulating objects in virtual environments.

We also investigated how head movements can serve to change the viewpoint in 3D applications, especially when the viewpoint needs to be changed quickly and temporarily to disambiguate the view. We studied how to use yaw and roll head movements to perform orbital camera control, i.e., to rotate the camera around a specific point in the scene [33]. We reported on four user studies. Study 1 evaluated the useful resolution of head movements and study 2 informed about visual and physical comfort. Study 3 compared two interaction techniques, designed by taking into account the results of the two previous studies. Results show that head roll is more efficient than head yaw for orbital camera control when interacting with a screen. Finally, Study 4 compared head roll with a standard technique relying on the mouse and the keyboard. Moreover, users were allowed to use both techniques at their convenience in a second stage. Results show that users prefer and are faster (14.5%) with the head control technique.

7.4.2. Storyboard sketching for stereo 3D films and Virtual Reality stories

The resurgence of stereoscopic and Virtual Reality (VR) media has motivated filmmakers to evolve new stereo- and VR-cinematic vocabularies, as many principles for stereo 3D film and VR story are unique. Concepts like plane separation, parallax position, and depth budgets in stereo, and presence, active experience, blocking and stitching in VR are missing from early planning due to the 2D nature of existing storyboards. Motivated to foresee difficulties exclusive to stereoscopy and VR, but also to exploit the unique possibilities of these medium, the 3D and VR cinematography communities encourages filmmakers to start thinking in stereo/VR as early as possible. Yet, there are very few early stage tools to support the ideation and discussion of a stereoscopic film or a VR story. Traditional solutions for early visual development and design, in current practices, are either strictly 2D or require 3D modeling skills, producing content that is consumed passively by the creative team.

To fill the gap in the filmmakers' toolkit, we proposed *Storeoboard* [31], a system for stereo-cinematic conceptualization, via storyboard sketching directly in stereo (Figure 6); and a novel multi-device system supporting the planning of virtual reality stories. Our tools are the first of their kind, allowing filmmakers to explore, experiment and conceptualize ideas in stereo or VR early in the film pipeline, develop new stereo- and VR-cinematic constructs and foresee potential difficulties. Our solutions are the design outcome of interviews and field work with directors, stereographers, storyboard artists and VR professionals. Our core contributions are thus: 1) a principled approach to the design and development of the first stereoscopic storyboard system that allows the director and artists to explore both the stereoscopic space and concepts in real-time, addressing key HCI challenges tied to sketching in stereoscopy; and 2) a principled survey of the state of the art in cinematic VR planning to design the first multi-device system that supports a storyboard workflow for VR film. We evaluated our tools with focus group and individual user studies with storyboard artists and industry professionals. In [31], we also report on feedback from the director of a live action, feature film on which Storeoboard was deployed. Results suggest that our approaches provide the speed and functionality needed for early stage planning, and the artifacts to properly discuss steroscopic and VR films.

7.4.3. Tactile displays and vibrotactile feedback

Tactile displays have predominantly been used for information transfer using patterns or as assistive feedback for interactions. With recent advances in hardware for conveying increasingly rich tactile information that mirrors visual information, and the increasing viability of wearables that remain in constant contact with the skin, there is a compelling argument for exploring tactile interactions as rich as visual displays. As Direct Manipulation underlies much of the advances in visual interactions, we introduced *Direct Manipulation-enabled Tactile display* [29]. We defined the concepts of a tactile screen, tactile pixel, tactile pointer, and tactile target which enable tactile pointing, selection and drag & drop. We built a proof of concept tactile display and studied its precision limits. We further developed a performance model for DMTs based on a tactile target acquisition study, and studied user performance in a real-world DMT menu application. The results show that users are able to use the application with relative ease and speed.

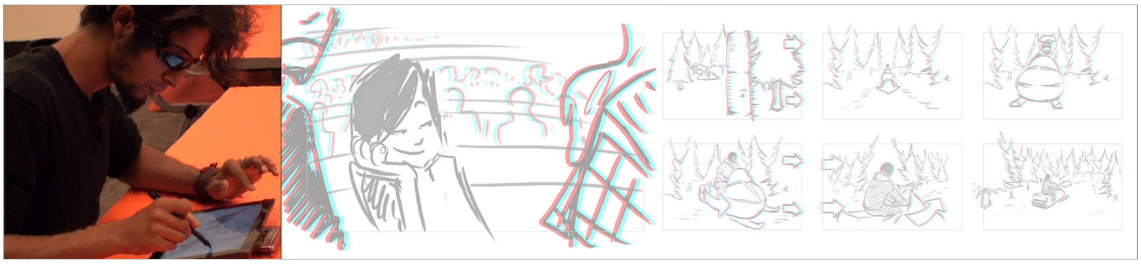


Figure 6. Storeoboard augments sketch-based storyboards with stereoscopic 3D planes for a fluid and flexible authoring of stereoscopic storyboards.

We have also explored vibrotactile feedback with wearable devices such as smartwatches and activity trackers, which are becoming prevalent. These devices provide continuous information about health and fitness, and offer personalized progress monitoring, often through multimodal feedback with embedded visual, audio, and vibrotactile displays. Vibrations are particularly useful when providing discreet feedback, without users having to look at a display or anyone else noticing, thus preserving the flow of the primary activity. Yet, current use of vibrations is limited to basic patterns, since representing more complex information with a single actuator is challenging. Moreover, it is unclear how much the user's current physical activity may interfere with their understanding of the vibrations. We addressed both issues through the design and evaluation of ActiVibe, a set of vibrotactile icons designed to represent progress through the values 1 to 10 [24]. We demonstrate a recognition rate of over 96% in a laboratory setting using a commercial smartwatch. ActiVibe was also evaluated in situ with 22 participants for a 28-day period. We show that the recognition rate is 88.7% in the wild and give a list of factors that affect the recognition, as well as provide design guidelines for communicating progress via vibrations.

7.4.4. Force-based autoscroll

Autoscroll, also known as edge-scrolling, is a common interaction technique in graphical interfaces that allows users to scroll a viewport while in dragging mode: once in dragging mode, the user moves the pointer near the viewport's edge to trigger an "automatic" scrolling. In spite of their wide use, existing autoscroll methods suffer from several limitations [45]. First, most autoscroll methods over-rely on the size of the control area, that is, the larger it is, the faster scrolling rate can be. Therefore, the level of control depends on the available distance between the viewport and the edge of the display, which can be limited. This is for example the case with small displays or when the view is maximized. Second, depending on the task, the users' intention can be ambiguous (e.g. dragging and dropping a file is ambiguous as the user's target may be located within the initial viewport or in a different one on the same display). To reduce this ambiguity, the size of the control area is drastically smaller for drag-and-drop operations which consequently also affects scrolling rate control as the user has a limited input area to control the scrolling speed.

We explored how force-sensing input, which is now available on commercial devices such as the Apple Magic Trackpad 2 or iPhone 6S, can be used to overcome the limitations of autoscroll. Indeed, force-sensing is an interesting candidate because: 1) users are often already applying a (relatively soft) force on the input device when using autoscroll and 2) varying force on the input device does not require to move the pointer, thus making it possible to offer control to the user while using a small and consistent control area regardless of the task and the device. We designed and proposed ForceEdge, a novel interaction technique mapping the force applied on a trackpad to the autoscrolling rate [19]. We implemented a software interface that can be used to design different transfer functions that map the force to autoscrolling rate and test these mappings for text selection and drag-and-drop tasks. Our pilot studies showed encouraging results and future work will focus on conducting more robust evaluations, as well as testing ForceEdge on mobile devices.

7.4.5. Combined Brain and gaze inputs for target selection

Gaze-based interfaces and Brain-Computer Interfaces (BCIs) allow for hands-free human–computer interaction. We investigated the combination of gaze and BCIs and proposed a novel selection technique for 2D target acquisition based on input fusion. This new approach combines the probabilistic models for each input, in order to better estimate the intent of the user. We evaluated its performance against the existing gaze and brain–computer interaction techniques. Twelve participants took part in our study, in which they had to search and select 2D targets with each of the evaluated techniques. Our fusion-based hybrid interaction technique was found to be more reliable than the previous gaze and BCI hybrid interaction techniques for 10 participants over 12, while being 29% faster on average. However, similarly to what has been observed in hybrid gaze-and-speech interaction, gaze-only interaction technique still provides the best performance. Our results should encourage the use of input fusion, as opposed to sequential interaction, in order to design better hybrid interfaces [14].

7.4.6. Actuated desktop devices

Desktop workstation remains the most common setup for office work tasks such as text editing, CAD, data analysis or programming. While several studies investigated how users interact with their devices (e.g. pressing keyboard keys, moving the cursor, etc.), it is not clear how they arrange their devices on the desk and whether we can leverage existing users' behaviors.

We designed the LivingDesktop [22], an augmented desktop with devices capable of moving autonomously. The LivingDesktop can control the position and orientation of the mouse, keyboard and monitors, offering different degrees of control for both the system (autonomous, semi-autonomous) and the user (manual, semi-manual) as well as different perceptive qualities (visual, haptic) thanks to a large range of device motions. We implemented a proof-of-concept of the LivingDesktop combining rail, robotic base and magnetism to control the position and orientation of the devices. This new setup presents several interesting features: (1) it improves ergonomics by continuously adjusting the position of its devices to help users adopting ergonomic postures and avoiding static postures for extended periods; (2) it facilitates collaborative works between local (e.g. located in the office) and remote co-workers; (3) it leverages context by reacting to the position of the user in the office, the presence of physical objects (e.g. tablets, food) or users' current activity in order to maintain a high level of comfort; (4) it reinforces physicality within the desktop workstation to increase immersion.

We conducted a scenario evaluation of the LivingDesktop. Our results showed the perceived usefulness of collaborative and ergonomics applications, as well as how it inspired our participants to elaborate novel application scenario, including social communication or accessibility.

7.4.7. Latency compensation

Human-computer interactions are greatly affected by the latency between the human input and the system visual response and the compensation of this latency is an important problem for the HCI community. We have developed a simple forecasting algorithm for latency compensation in indirect interaction using a mouse, based on numerical differentiation. Several differentiators were compared, including a novel algebraic version, and an optimized procedure was developed for tuning the parameters of the algorithm. The efficiency was demonstrated on real data, measured with a 1ms sampling time. These results are developed in [37] and patent has been filed on a subsequent technique for latency compensation [42].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- **Mock-up of a tool for dynamic media pre-production:** we are currently working with the HCOP holding company on the design of new tools for the pre-production of dynamic medias such as videos, e-learning animations, etc. This work involves interviews of professional video producers, the identification of opportunities for tools that could help them, and the production of descriptions and mock-ups of these tools.

- **Recognition and interpretation of piano fingering:** we have started a new collaboration with Hugues Leclère, concert pianist and professor at the "Conservatoire à rayonnement régional de Paris". Our objective is to investigate new sensing technology and interpretation algorithms for accurate live recognition of piano fingerings. Ultimately, this technology would ease the transcription of fingerings directly onto scores during play and support both the learning and training of piano fingerings, given appropriate visualization and interaction techniques that we will investigate in a second phase of this collaboration.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Mjolnir/UCLIC associate team (Inria Lille, 2015-2017)

Participants: Sylvain Malacria [correspondent], Nicolas Roussel.

The goal of this project is the design and implementation of novel cross-device systems and interaction techniques that minimize the cost of divided attention. Of particular interest are notification systems on smart watches and in distributed computing systems. More precisely, we design cross-device activity and notifications monitor that will intercept external (e.g. new e-mail) and internal (e.g. a video editing software completed an export) notifications and distribute them to the device users are currently wearing/interacting with in order to minimize notification redundancy.

Partner: University College London Interaction Centre (United Kingdom).

9.2. National Initiatives

9.2.1. Turbotouch (ANR, 2014-2018)

Participants: Géry Casiez [correspondent], Nicolas Roussel, Thomas Pietrzak.

Touch-based interactions with computing systems are greatly affected by two interrelated factors: the transfer functions applied on finger movements, and latency. This project aims at transforming the design of touch transfer functions from black art to science to support high-performance interactions. We are working on the precise characterization of the functions used and the latency observed in current touch systems. We are developing a testbed environment to support multidisciplinary research on touch transfer functions and will use this testbed to design latency reduction and compensation techniques, and new transfer functions.

Partners: Inria Lille's NON-A team and the "Perceptual-motor behavior group" from the Institute of Movement Sciences.

Web site: http://mjolnir.lille.inria.fr/turbotouch/

Related publications: [20], [35], [19], [34], [37], [42]

9.2.2. ParkEvolution (Carnot Inria - Carnot STAR, 2015-2017)

Participants: Géry Casiez [correspondent], Sébastien Poulmane.

This project studies the fine motor control of patients with Parkinson disease in an ecological environment, at home, without the presence of experimenters. Through longitudinal studies, we collect raw information from pointing devices to create a large database of pointing behavior data. From the analysis of this big dataset, the project aims at inferring the individual's disease progression and influence of treatments.

Partners: the "Perceptual-motor behavior group" from the Institute of Movement Sciences and Hôpital de la Timone.

Web site: http://parkevolution.org/

9.2.3. BCI-LIFT (Inria Project Lab, 2015-2019)

Participants: Géry Casiez, Nicolas Roussel [correspondent].

The goal of this large-scale initiative is to design a new generation of non-invasive Brain-Computer Interfaces (BCI) that are easier to appropriate, more efficient, and suited for a larger number of people.

Partners: Inria's ATHENA, NEUROSYS, POTIOC, HYBRID & DEMAR teams, *Centre de Recherche en Neurosciences de Lyon* (INSERM) and INSA Rouen.

Web site: https://bci-lift.inria.fr/

Related publications: [38], [15], [14], [41]

9.3. European Initiatives

9.3.1. Happiness (H2020-ICT-2014-1/ICT-03-2014/RIA, 2015-2018)

Participants: Christian Frisson, Julien Decaudin, Thomas Pietrzak [correspondent], Nicolas Roussel.

The main objective of this project is to develop and evaluate new types of haptic actuators printed on advanced Thin, Organic and Large Area Electronics (TOLAE) technologies for use in car dashboards. These actuators are embedded in plastic molded dashboard parts. The expected outcome is a marketable solution for haptic feedback on curved interactive surfaces.

Partners: CEA (coordinator), Inria Rennes' HYBRID team, Arkema, Bosch, Glasgow University, ISD, Walter Pack, Fundacion Gaiker.

Web site: http://happiness-project.eu/

Related publication: [29]

9.4. International Initiatives

9.4.1. MIDWAY (Inria associate team, 2014-2016)

Participants: Fanny Chevalier, Stéphane Huot [correspondent], Justin Mathew.

The goal of the project is the design and implementation of a musical interaction design workbench to facilitate the exploration and definition of new interactive technologies for both musical creation and performance.

Partner: Inria Saclay's EXSITU team and the Input Devices and Music Interaction Laboratory (IDMIL) from the Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT) at McGill University, Canada.

Web site: http://insitu.lri.fr/MIDWAY/

Related publication: [43]

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Visiting scholars:

- Marcelo Wanderley, Professor at McGill University, Canada (3 one week visits in April, October & December)
- Edward Lank, Associate Professor at the University of Waterloo, Canada (since September)
- Daniel Wigdor, Associate Professor at the University of Toronto, Canada (April 2016)
- Baptiste Caramiaux, Post-Doctoral researcher at McGill University, Canada, & IRCAM (December)

Internships:

- Filipe Calegario, PhD student at McGill University, Canada (January)
- Nicholas Fellion, Master's student at Carleton University, Canada (from January to April)
- Aakar Gupta, PhD student at the University of Toronto, Canada (from June to September)
- Hrim Mehta, PhD student at the Ontario Institute of Technology, Canada (from May to August)
- Anastasia Kuzminykh, PhD student at the University of Waterloo, Canada (from October to December)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- VISAP (IEEE): Fanny Chevalier (general co-chair)
- Rencontre Inria Industrie "Interactions avec les objets et services numériques": Nicolas Roussel (scientific chair)

10.1.1.2. Member of the Organizing Committees

- IHM: Fanny Chevalier (Work-in-Progress co-chair), Stéphane Huot (AFIHM's 20th anniversary co-chair)

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- CHI (ACM): Géry Casiez (associate chair), Nicolas Roussel (associate chair)
- UIST (ACM): Fanny Chevalier (associate chair), Géry Casiez (associate chair)
- Infovis (IEEE): Fanny Chevalier (PC member)
- MobileHCI (ACM): Sylvain Malacria (associate chair)
- IHM: Fanny Chevalier (PC member)

10.1.2.2. Reviewer

- SIGGRAPH (ACM): Fanny Chevalier
- UIST (ACM): Sylvain Malacria, Thomas Pietrzak
- IHM: Sylvain Malacria
- GI (ACM): Sylvain Malacria
- TEI (ACM): Thomas Pietrzak
- MobileHCI (ACM): Thomas Pietrzak
- Haptic Symposium (IEEE): Thomas Pietrzak

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- *Transactions on Computer Human Interaction* (ACM): Sylvain Malacria, Géry Casiez, Thomas Pietrzak
- *Transactions on Visualization and Computer Graphics* (IEEE): Fanny Chevalier
- *Information Design Journal* (John Benjamins Publishing Company): Fanny Chevalier
- *Interacting with Computers* (Oxford Journals): Thomas Pietrzak

10.1.4. Invited Talks

- *Direct spacetime sketching and editing of visual media*, Microsoft Research, Redmond: Fanny Chevalier
- *Opportunities and limits of Adaptive User Interfaces*, Technology and Routines workshop, Bavarian Academy of Sciences and Humanities, Munich, Germany: Sylvain Malacria
- *Towards the full experience of playing drums on a virtual drumkit*, Symposium on Haptic and Music, McGill University, Montréal, Canada: Thomas Pietrzak

10.1.5. Scientific Expertise

- PICOM competitiveness cluster: Nicolas Roussel (scientific board member)
- Agence Nationale de la Recherche: Fanny Chevalier (committee member for the *JCJC*, *PRCE* and *PRC* programs)
- European Research Council: Géry Casiez (reviewer for the Starting Grant program)

10.1.6. Research Administration

For Inria:

- Evaluation committee: Nicolas Roussel (member)
- Gender equity and equality committee: Nicolas Roussel (member)
- Strategic orientation committee for the information system (COSS scientifique): Nicolas Roussel (member)
- International relations working group (COST-GTRI): Stéphane Huot (member)

For Inria Lille – Nord Europe:

- Scientific officer (délégué scientifique): Nicolas Roussel
- Research jobs committee (CER): Fanny Chevalier (member), Nicolas Roussel (member)
- Technical development committee (CDT): Nicolas Roussel (member)
- Consultative committee (Comité de centre): Fanny Chevalier (member)
- Operational legal and ethical risk assessment committee (COERLE): Stéphane Huot (local correspondent)
- Support to researchers (accompagnement des chercheurs): Stéphane Huot (adviser)
- Activity reports (RAweb): Nicolas Roussel (local correspondents)

For the CRIStAL lab of Univ. Lille 1:

- Laboratory council: Géry Casiez (elected member)

10.1.7. Learned societies

- Association Francophone d'Interaction Homme-Machine (AFIHM): Géry Casiez (board member, president), Stéphane Huot (board member and scientific council member), Thomas Pietrzak (board member)

10.1.8. Hiring committees

- Inria's eligibility jury for senior researcher positions (DR2): Nicolas Roussel
- Inria's eligibility jury for junior researcher positions (CR2) in Saclay: Nicolas Roussel
- ENAC's hiring committee for an assistant professor position: Nicolas Roussel
- Polytech Nantes' hiring committee for a Computer Science Assistant Professor position: Fanny Chevalier (member)

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- Master Informatique *Image Vision Interaction (IVI)*: Géry Casiez (16 hrs), Fanny Chevalier (16 hrs), Thomas Pietrzak (16 hrs), *NIHM : Nouvelles Interactions Homme-Machine*, M2, Univ. Lille 1
- Master Informatique: Thomas Pietrzak (50 hrs), Sylvain Malacria (30 hrs), Nicolas Roussel (8 hrs), *IHM : Interactions Homme-Machine*, M1, Univ. Lille 1
- Master Informatique: Géry Casiez, *Multi-Touch Interaction*, 10 hrs, M1, Univ. Lille 1
- Master: Géry Casiez (6 hrs), Thomas Pietrzak (10.5 hrs), *3DETech : 3D Digital Entertainment Technologies*, M2 level, Télécom Lille
- Master Sciences Humaines et Sociales *Réseaux Sociaux Numériques (RSN)*: Fanny Chevalier, *Statistiques et visualisation de données*, 8hrs, M2, Univ. Lille 1
- Master Sciences Humaines et Sociales *Réseaux Sociaux Numériques (RSN)*: Fanny Chevalier, *Outils et pratiques numériques*, 6hrs, M1, Univ. Lille 1
- Master Computing for Medicine: Fanny Chevalier, *Information Visualization in Medicine*, 1hr, Masters, Univ. of Toronto
- Graduate course in Computer Sciences: Fanny Chevalier, *Topics in Interactive Computing: Information Visualization*, 16hrs, MSc and PhD, Univ. of Toronto
- Licence Informatique: Thomas Pietrzak (45 hrs), *Logique*, L3, Univ. Lille 1
- Licence Informatique: Thomas Pietrzak (36 hrs), *Automates et Langages*, L3, Univ. Lille 1
- Licence Sciences pour l'Ingénieur (SPI): Sylvain Malacria, *Introduction à l'Interaction Homme Machine*, 26hrs, L3, Institut Villebon Georges Charpak
- Research Masters in Computer Science & HCID Masters (EIT ICT Labs European Master in Human-Computer Interaction and Design): Stéphane Huot, *Advanced Programming of Interactive Systems*, 13.5 hrs, M1 & M2, Univ. Paris Saclay
- DUT Informatique: Géry Casiez, *IHM*, 72 hrs, 1^{st} year, IUT A de Lille - Univ. Lille 1
- DUT Informatique: Géry Casiez, *Algorithmique*, 80 hrs, 1^{st} year, IUT A de Lille - Univ. Lille 1
- DUT Informatique: Géry Casiez, *Modélisation mathématique*, 14 hrs, 2^{nd} year, IUT A de Lille - Univ. Lille 1
- DUT Informatique: Géry Casiez, *Projets*, 18 hrs, 2^{nd} year, IUT A de Lille - Univ. Lille 1

10.2.2. Supervision

PhD in progress: Hakim Si Mohammed, *Improving Interaction based on a Brain-Computer Interface*, started October 2016, advised by Anatole Lecuyer, Géry Casiez & Nicolas Roussel (in Rennes)

PhD in progress: Nicole Ke Chen Pong, *Comprendre et améliorer le vocabulaire interactionel des utilisateurs*, started October 2016, advised by Nicolas Roussel & Sylvain Malacria

PhD in progress: Thibault Raffaillac, *Languages and system infrastructure for interaction*, started October 2015, advised by Stéphane Huot & Stéphane Ducasse

PhD in progress: Amira Chalbi-Neffati, *Comprendre et mieux concevoir les animations dans le contexte des interfaces graphiques*, started October 2014, advised by Nicolas Roussel & Fanny Chevalier

PhD in progress: Jeronimo Barbosa, *Design and Evaluation of Digital Musical Instruments*, McGill University, started in 2013, advised by Marcelo Wanderley & Stéphane Huot (since 2016)

PhD in progress: Alexandre Kouyoumdjian, *Multimodal selection of numerous moving targets in large visualization platforms: application to interactive molecular simulation*, started October 2013, advised by Stéphane Huot, Patrick Bourdot & Nicolas Ferey (in Saclay)

PhD in progress: Justin Mathew, *New visualization and interaction techniques for 3D spatial audio*, started June 2013, advised by Stéphane Huot & Brian Katz (in Saclay)

PhD: Oleksandr Zinenko, *Interactive code restructuring*, Univ. Paris Saclay, defended in November 2016, advised by Stéphane Huot & Cédric Bastoul (Université de Strasbourg)

PhD: Alix Goguey, *Understanding and designing touch interaction using finger identification*, Univ. Lille 1, defended in October 2016 [11], advised by Géry Casiez.

PhD: Andéol Evain, *Optimizing the use of SSVEP-based brain-computer interfaces for human-computer interaction*, defended in December 2016, advised by Anatole Lecuyer, Géry Casiez & Nicolas Roussel (in Rennes)

10.2.3. Thesis committees

- Michael Glueck (PhD, University of Toronto): Fanny Chevalier
- Nicole Sultanum (PhD, University of Toronto): Fanny Chevalier
- Ignacio Avellino (PhD, Université Paris Saclay): Nicolas Roussel

10.2.4. Juries

- Morten Esbensen (PhD, IT University of Copenhagen): Nicolas Roussel, reviewer
- Élisabeth Rousset (PhD, Université de Grenoble): Géry Casiez, reviewer
- Sébastien Pelurson (PhD, Université de Grenoble): Géry Casiez, reviewer
- Waseem Safi (PhD, Université Cean Normandie): Thomas Pietrzak, examiner

10.3. Popularization

- *"Vers une meilleure appréciation des algorithmes qui nous entourent"*, talk for Rencontres Inria Industrie, Plaine Image, Tourcoing, November: Fanny Chevalier
- Prospective talk on HCI as part of BeyondLab's "*Aux portes du futur*" evening, November: Nicolas Roussel
- *"Chorégraphie des Transitions animées"*, talk for CRIStAL journée Recherche Innovation Création, Lille, October: Amira Chalbi-Neffati
- "*Les objets deviennent intelligents, et nous ?*", talk at Lycée Diderot (Carvin) as part of Inria Lille's "Chercheurs Itinérants" program, October: Nicolas Roussel
- *"Efficacité et performance des transitions animées"*, talk at EuraTechnologies (plateau Inria), March: Fanny Chevalier
- Presentation of the Mjolnir team, talk at EuraTechnologies (plateau Inria), January: Nicolas Roussel

11. Bibliography

Major publications by the team in recent years

[1] G. BAILLY, T. PIETRZAK, J. DEBER, D. J. WIGDOR.*Metamorphe: augmenting hotkey usage with actuated keys*, in "Proceedings of CHI'13", ACM, April 2013, p. 563-572, http://dx.doi.org/10.1145/2470654.2470734.

[2] G. CASIEZ, S. CONVERSY, M. FALCE, S. HUOT, N. ROUSSEL.*Looking through the eye of the mouse: a simple method for measuring end-to-end latency using an optical mouse*, in "Proceedings of UIST'15", ACM, November 2015, p. 629-636, http://dx.doi.org/10.1145/2807442.2807454.

[3] G. CASIEZ, N. ROUSSEL.*No more bricolage! Methods and tools to characterize, replicate and compare pointing transfer functions*, in "Proceedings of UIST'11", ACM, October 2011, p. 603-614, http://dx.doi.org/10.1145/2047196.2047276.

[4] G. CASIEZ, N. ROUSSEL, D. VOGEL.*1 euro filter: a simple speed-based low-pass filter for noisy input in interactive systems*, in "Proceedings of CHI'12", ACM, May 2012, p. 2527-2530, http://doi.acm.org/10.1145/2207676.2208639.

[5] A. COCKBURN, C. GUTWIN, J. SCARR, S. MALACRIA.*Supporting novice to expert transitions in user interfaces*, in "ACM Computing Surveys", November 2014, vol. 47, n^{o} 2, http://dx.doi.org/10.1145/2659796.

[6] A. GUPTA, T. PIETRZAK, N. ROUSSEL, R. BALAKRISHNAN.*Direct manipulation in tactile displays*, in "Proceedings of CHI'16", ACM, May 2016, p. 3683-3693, http://dx.doi.org/10.1145/2858036.2858161.

[7] R. H. KAZI, F. CHEVALIER, T. GROSSMAN, S. ZHAO, G. FITZMAURICE.*DRACO: bringing life to illustrations with kinetic textures*, in "Proceedings of CHI'14", ACM, April 2014, p. 351-360, http://dx.doi.org/10.1145/2556288.2556987.

[8] S. MALACRIA, G. BAILLY, J. HARRISON, A. COCKBURN, C. GUTWIN.*Promoting hotkey use through rehearsal with ExposeHK*, in "Proceedings of CHI'13", ACM, April 2013, p. 573-582, http://doi.acm.org/10.1145/2470654.2470735.

[9] J. WAGNER, S. HUOT, W. E. MACKAY.*BiTouch and BiPad: designing bimanual interaction for hand-held tablets*, in "Proceedings of CHI'12", ACM, 2012, p. 2317-2326, http://doi.acm.org/10.1145/2207676.2208391.

[10] J. ZHAO, M. GLUECK, F. CHEVALIER, Y. WU, A. KHAN.*Egocentric analysis of dynamic networks with EgoLines*, in "Proceedings of CHI'16", ACM, May 2016, p. 5003-5014, http://dx.doi.org/10.1145/2858036.2858488.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. GOGUEY.*Understanding and designing touch interaction using finger identification*, Université Lille 1 - Sciences et Technologies, October 2016, https://tel.archives-ouvertes.fr/tel-01390458.

Articles in International Peer-Reviewed Journal

[12] M. GLUECK, A. GVOZDIK, F. CHEVALIER, A. KHAN, M. BRUDNO, D. J. WIGDOR.*PhenoStacks: Cross-Sectional Cohort Phenotype Comparison Visualizations*, in "IEEE Transactions on Visualization and Computer Graphics", October 2016 [*DOI :* 10.1109/TVCG.2016.2598469], https://hal.inria.fr/hal-01353233.

[13] J. ZHAO, M. GLUECK, S. BRESLAV, F. CHEVALIER, A. KHAN.*Annotation Graphs: A Graph-Based Visualization for Meta-Analysis of Data based on User-Authored Annotations*, in "IEEE Transactions on Visualization and Computer Graphics", October 2016 [*DOI :* 10.1109/TVCG.2016.2598543], https://hal.inria.fr/hal-01353368.

[14] A. ÉVAIN, F. ARGELAGUET, G. CASIEZ, N. ROUSSEL, A. LÉCUYER.*Design and Evaluation of Fusion Approach for Combining Brain and Gaze Inputs for Target Selection*, in "Frontiers in Neuroscience", October 2016, vol. 10, n^{o} 454, 14 [*DOI :* 10.3389/FNINS.2016.00454], https://hal.inria.fr/hal-01388528.

[15] A. ÉVAIN, F. ARGELAGUET, G. CASIEZ, N. ROUSSEL, A. LÉCUYER.*Do the stimuli of an SSVEP-based BCI really have to be the same as the stimuli used for training it?*, in "Brain-Computer Interfaces", July 2016, vol. 3, n^{o} 2, p. 103 - 111 [*DOI :* 10.1080/2326263X.2016.1193458], https://hal.inria.fr/hal-01388534.

Articles in Non Peer-Reviewed Journal

[16] T. GROSSMAN, F. CHEVALIER, R. H. KAZI.*Bringing research articles to life with animated figures*, in "Interactions", August 2016, vol. 23, n^{o} 4, p. 52-57 [*DOI :* 10.1145/2949762], https://hal.inria.fr/hal-01352860.

Invited Conferences

[17] T. PIETRZAK.*Towards the Full Experience of Playing Drums on a Virtual Drum Kit*, in "Symposium on Force Feedback and Music", Montréal, Canada, December 2016, 1, https://hal.inria.fr/hal-01417872.

International Conferences with Proceedings

[18] M. ACHIBET, G. CASIEZ, M. MARCHAL.*DesktopGlove: a Multi-finger Force Feedback Interface Separating Degrees of Freedom Between Hands*, in "3DUI'16, the 11th Symposium on 3D User Interfaces", Greenville, United States, I. C. SOCIETY (editor), March 2016, p. 3-12 [*DOI :* 10.1109/3DUI.2016.7460024], https://hal.inria.fr/hal-01267645.

[19] A. ANTOINE, S. MALACRIA, G. CASIEZ.*Controlling autoscroll on trackpads using the force*, in "Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine", Fribourg, Switzerland, Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine, October 2016, p. 264-270 [*DOI :* 10.1145/3004107.3004137], https://hal.archives-ouvertes.fr/hal-01384315.

[20] S. ARANOVSKIY, R. USHIROBIRA, D. EFIMOV, G. CASIEZ.*Modeling Pointing Tasks in Mouse-Based Human-Computer Interactions*, in "CDC 2016 - 55th IEEE Conference on Decision and Control", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01380318.

[21] L. BAGNÈRES, O. ZINENKO, S. HUOT, C. BASTOUL.*Opening Polyhedral Compiler's Black Box*, in "CGO 2016 - 14th Annual IEEE/ACM International Symposium on Code Generation and Optimization", Barcelona, Spain, March 2016, https://hal.inria.fr/hal-01253322.

[22] G. BAILLY, S. SAHDEV, S. MALACRIA, T. PIETRZAK.*LivingDesktop: Augmenting Desktop Workstation with Actuated Devices*, in "Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)", San Jose, CA, United States, May 2016, p. 5298 - 5310 [*DOI :* 10.1145/2858036.2858208], https://hal.archives-ouvertes.fr/hal-01403719.

[23] *Best Paper*
N. BANOVIC, T. BUZALI, F. CHEVALIER, J. MANKOFF, A. K. DEY.*Modeling and Understanding Human Routine Behavior*, in "ACM CHI Conference on Human Factors in Computing Systems 2016", Santa Clara, California, United States, ACM, May 2016, p. 248 - 260 [*DOI :* 10.1145/2858036.2858557], https://hal.inria.fr/hal-01416119.

[24] J. R. CAUCHARD, J. L. CHENG, T. PIETRZAK, J. A. LANDAY.*ActiVibe: Design and Evaluation of Vibrations for Progress Monitoring*, in "CHI '16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems", San Jose, United States, May 2016 [*DOI :* 10.1145/2858036.2858046], https://hal.inria.fr/hal-01417862.

[25] F. CHEVALIER, N. HENRY RICHE, C. PLAISANT, A. CHALBI, C. HURTER.*Animations 25 Years Later: New Roles and Opportunities*, in "AVI 16, International Working Conference on Advanced Visual Interfaces", Bari, Italy, AVI '16 Proceedings of the International Working Conference on Advanced Visual Interfaces, ACM, June 2016, p. 280-287 / ISBN : 978-1-4503-4131-8 [*DOI :* 10.1145/2909132.2909255], https://hal-enac.archives-ouvertes.fr/hal-01322962.

[26] C. FRISSON, S. MALACRIA, G. BAILLY, D. THIERRY.*InspectorWidget: a System to Analyze Users Behaviors in Their Applications*, in "CHI'16 Extended Abstracts", San Jose, United States, May 2016 [*DOI :* 10.1145/2851581.2892388], https://hal.inria.fr/hal-01418184.

[27] C. FRISSON, T. PIETRZAK, S. ZHAO, Z. SCHWEMLER, A. ISRAR.*WebAudioHaptics: Tutorial on Haptics with Web Audio*, in "Web Audio Conference (WAC)", Atlanta, United States, Web Audio Conference (WAC), April 2016, https://hal.inria.fr/hal-01429161.

[28] A. GOGUEY, M. NANCEL, G. CASIEZ, D. VOGEL.*The Performance and Preference of Different Fingers and Chords for Pointing, Dragging, and Object Transformation*, in "In Proceedings of CHI'16, the 34th Conference on Human Factors in Computing Systems", San Jose, United States, ACM, May 2016, p. 4250-4261 [*DOI :* 10.1145/2858036.2858194], https://hal.inria.fr/hal-01267607.

[29] *Best Paper*
A. GUPTA, T. PIETRZAK, N. ROUSSEL, R. BALAKRISHNAN.*Direct Manipulation in Tactile Displays*, in "CHI 2016 - Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems", San Jose, United States, May 2016, p. 3683 - 3693 [*DOI :* 10.1145/2858036.2858161], https://hal.inria.fr/hal-01417855.

[30] R. HENRIKSON, B. DE ARAUJO, F. CHEVALIER, K. SINGH, R. BALAKRISHNAN.*Multi-Device Storyboards for Cinematic Narratives in VR*, in "ACM Symposium on User Interface Software and Technology (UIST '16)", Tokyo, Japan, ACM, October 2016, p. 787-796 [*DOI :* 10.1145/2984511.2984539], https://hal.inria.fr/hal-01416153.

[31] R. HENRIKSON, B. DE ARAUJO, F. CHEVALIER, K. SINGH, R. BALAKRISHNAN.*Storeoboard: Sketching Stereoscopic Storyboards*, in "ACM CHI Conference on Human Factors in Computing Systems 2016", Santa Clara, California, United States, May 2016, p. 4587 - 4598 [*DOI :* 10.1145/2858036.2858079], https://hal.inria.fr/hal-01416137.

[32] A. ISRAR, S. ZHAO, K. MCINTOSH, Z. SCHWEMLER, A. FRITZ, J. MARS, J. BEDFORD, C. FRISSON, I. HUERTA, M. KOSEK, B. KONIARIS, K. MITCHELL.*Stereohaptics: A Haptic Interaction Toolkit for Tangible Virtual Experiences*, in "ACM SIGGRAPH 2016 Studio", Anaheim, United States, July 2016 [*DOI :* 10.1145/2929484.2970273], https://hal.inria.fr/hal-01429160.

[33] T. JACOB, G. BAILLY, L. ERIC, G. CASIEZ, M. TEYSSIER.*Desktop Orbital Camera Motions Using Rotational Head Movements*, in "In proceedings of SUI'16, the ACM Symposium on Spatial User Interac-

tion", Tokyo, Japan, October 2016, p. 139-148 [*DOI :* 10.1145/2983310.2985758], https://hal.inria.fr/hal-01420664.

[34] S. MALACRIA, A. GOGUEY, G. BAILLY, G. CASIEZ.*Multi-touch trackpads in the wild*, in "Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine", Fribourg, Switzerland, Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine, October 2016, p. 19-24 [*DOI :* 10.1145/3004107.3004113], https://hal.archives-ouvertes.fr/hal-01383528.

[35] M. NANCEL, D. VOGEL, B. DE ARAUJO, R. JOTA, G. CASIEZ.*Next-Point Prediction Metrics for Perceived Spatial Errors*, in "In proceedings of UIST'16, the 29th ACM Symposium on User Interface Software and Technology", Tokyo, Japan, October 2016, p. 271-285 [*DOI :* 10.1145/2984511.2984590], https://hal.inria.fr/hal-01420670.

[36] L. POTIER, T. PIETRZAK, G. CASIEZ, N. ROUSSEL.*Designing tactile patterns with programmable friction*, in "Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine", Fribourg, Switzerland, Actes de la 28ième conférence francophone sur l'Interaction Homme-Machine, October 2016, p. 1-7 [*DOI :* 10.1145/3004107.3004110], https://hal.archives-ouvertes.fr/hal-01383516.

[37] R. USHIROBIRA, D. EFIMOV, G. CASIEZ, N. ROUSSEL, W. PERRUQUETTI.*A forecasting algorithm for latency compensation in indirect interaction*, in "In proceedings of ECC'16, the 15th annual European Control Conference", Aalborg, Denmark, June 2016, 6, https://hal.inria.fr/hal-01420653.

[38] A. ÉVAIN, F. ARGELAGUET SANZ, A. STROCK, N. ROUSSEL, G. CASIEZ, A. LÉCUYER.*Influence of Error Rate on Frustration of BCI Users*, in "AVI'16 - International Working Conference on Advanced Visual Interfaces", Bari, Italy, ACM, June 2016, p. 248–251 [*DOI :* 10.1145/2909132.2909278], https://hal.inria.fr/hal-01388552.

Conferences without Proceedings

[39] A. CHALBI.*Comprendre et Concevoir les Animations dans le Contexte des Interfaces Graphiques*, in "IHM 16 - 28ème Conférence Francophone sur l'Interaction Homme-Machine", Fribourg, Switzerland, October 2016, https://hal.inria.fr/hal-01417256.

[40] *Best Paper*
J. ZHAO, M. GLUECK, F. CHEVALIER, Y. WU, A. KHAN.*Egocentric Analysis of Dynamic Networks with EgoLines*, in "CHI '16 - Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems", San Jose, CA, United States, ACM, May 2016, p. 5003-5014 [*DOI :* 10.1145/2858036.2858488], https://hal.inria.fr/hal-01325788.

Scientific Books (or Scientific Book chapters)

[41] A. ÉVAIN, N. ROUSSEL, G. CASIEZ, F. ARGELAGUET SANZ, A. LÉCUYER.*Brain-Computer Interfaces for Human–Computer Interaction*, in "Brain-Computer Interfaces 1: Foundations and Methods", M. CLERC, L. BOUGRAIN, F. LOTTE (editors), Wiley-ISTE, August 2016, https://hal.inria.fr/hal-01416382.

Patents and standards

[42] S. ARANOVSKIY, G. CASIEZ, D. EFIMOV, N. ROUSSEL, R. USHIROBIRA.*Dispositif a` affichage prédictif*, September 2016, n^o 1000365810, https://hal.inria.fr/hal-01426069.

Other Publications

[43] M. M. WANDERLEY, J. MALLOCH, J. GARCIA, W. E. MACKAY, M. BEAUDOUIN-LAFON, S. HUOT.*Human Computer Interaction meets Computer Music: The MIDWAY Project*, May 2016, 4, Workshop Music and HCI, in conjunction with ACM Conference on Human Factors in Computing Systems (CHI '16), https://hal.inria.fr/hal-01370588.

[44] O. ZINENKO, S. HUOT, C. BASTOUL.*Interactive Program Restructuring*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01406294.

References in notes

[45] J. ACEITUNO.*Direct and expressive interaction on the desktop: increasing granularity, extent, and dimensionality*, Université Lille 1, France, Octobre 2015, 240 pages.

[46] G. CASIEZ, S. CONVERSY, M. FALCE, S. HUOT, N. ROUSSEL.*Looking through the eye of the mouse: a simple method for measuring end-to-end latency using an optical mouse*, in "Proceedings of UIST'15", ACM, November 2015, p. 629-636, http://dx.doi.org/10.1145/2807442.2807454.

[47] M. CHALMERS, A. GALANI.*Seamful interweaving: heterogeneity in the theory and design of interactive systems*, in "Proceedings of DIS'04", ACM, 2004, p. 243-252, http://doi.acm.org/10.1145/1013115.1013149.

[48] D. C. ENGELBART.*Augmenting human intellect: a conceptual framework*, Stanford Research Institute, October 1962, n° AFOSR-3233, http://www.dougengelbart.org/pubs/augment-3906.html.

[49] M. GLUECK, P. HAMILTON, F. CHEVALIER, S. BRESLAV, A. KHAN, D. WIGDOR, M. BRUDNO.*PhenoBlocks: phenotype comparison visualizations*, in "Proceedings of VAST'15", IEEE, October 2015, p. 1077-2626, http://dx.doi.org/10.1109/TVCG.2015.2467733.

[50] M. HEIDEGGER.*Sein und zeit*, Max Niemeyer Verlag, February 1927.

[51] E. PACHERIE.*The phenomenology of action: a conceptual framework*, in "Cognition", 2008, vol. 107, n° 1, p. 179 - 217, http://dx.doi.org/10.1016/j.cognition.2007.09.003.

Project-Team MODAL

MOdel for Data Analysis and Learning

IN COLLABORATION WITH: Laboratoire Paul Painlevé (LPP)

IN PARTNERSHIP WITH:
CNRS

Université Lille 2

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Optimization, machine learning and statistical methods

Table of contents

Project-Team MODAL

Creation of the Team: 2010 September 01, updated into Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

3.1.4. - Uncertain data
3.2.3. - Inference
3.3.2. - Data mining
3.3.3. - Big data analysis
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.5. - Bayesian methods
3.4.7. - Kernel methods
5.2. - Data visualization
6.2.3. - Probabilistic methods
6.2.4. - Statistical methods
6.3.3. - Data processing
8.2. - Machine learning

Other Research Topics and Application Domains:

1.1.6. - Genomics
2.2.3. - Cancer
9.4.5. - Data science
9.5.3. - Economy, Finance
9.5.5. - Sociology

1. Members

Research Scientist

Benjamin Guedj [Inria, Research Scientist, Researcher]

Faculty Members

Christophe Biernacki [Team leader, Univ. Lille I, Faculty Member, Professor, HDR]
Alain Celisse [Univ. Lille I, Faculty Member, Associate Professor]
Serge Iovleff [Univ. Lille I, Faculty Member, Associate Professor]
Guillemette Marot [Univ. Lille II, Faculty Member, Associate Professor]
Cristian Preda [Univ. Lille I, Faculty Member, Professor, HDR]
Vincent Vandewalle [Univ. Lille II, Faculty Member, Associate Professor]

Technical Staff

Quentin Grimonprez [PhD Student funded by Univ. Lille 1 until Sep 2016, then Inria Engineer funded by Région Hauts-de-France]
Etienne Goffinet [Inria, granted by Région Hauts-de-France]
Vincent Kubicki [Inria, granted by Région Hauts-de-France]
Matthieu Marbac Lourdelle [Inria, from Dec 2016]

PhD Students

Maxime Baelde [A-Volute, granted by CIFRE]

Anne-Lise Bedenel [PIXEO, granted by CIFRE]
Maxime Brunin [Univ. Lille I]
Adrien Ehrhardt [CACF, from Apr 2016, granted by CIFRE]
Jeremie Kellner [Univ. Lille I, until Sep 2016]
Le Li [iAdvize, granted by CIFRE]

Post-Doctoral Fellows

Alexandru Amarioarei [Inria, granted by ANR CLINMINE project]
Cristina Preda [Inria, until Aug 2016]

Visiting Scientist

Emmanuel Chazard [Univ. Lille II, until Jul 2016]

Administrative Assistant

Anne Rejl [Inria]

Others

Samuel Blanck [Univ. Lille II, External Member]
Miguel Assuncao [Inria, Intern, from Sep 2016]
Rohit Uttam Bhagwat [Inria, Intern, from Jun 2016 until Aug 2016]
Faicel Chamroukhi [Univ. Caen, External Member, from Sep 2016, HDR]
Olivier Delrieu [C4X Discovery, External Member]
Philippe Heinrich [Univ. Lille I, External Member, from Jul 2016]
Hamza Cherkaoui [Inria, Intern, from May 2016 until Aug 2016]
Sophie Dabo [Univ. Lille III, External Member, HDR]
Astha Gupta [Inria, Intern, from May 2016 until Jul 2016]
Julien Jacques [Univ. Lyon II, External Member, HDR]
Siddharth Sharma [Inria, Intern, until May 2016]
Bhargav Srinivasa Desikan [Inria, Intern, from Aug 2016]

2. Overall Objectives

2.1. Overall Objectives

Modal is a team focused on statistical methodology for data analysis (clustering, visualization) and learning (classification, density estimation, aggregation, matrix factorization, ...). In this context, the core of the team's work is to design meaningful generative models for prominent complex data (mixed structured data), which are still almost ignored in the literature. From those generative models, learning procedures are proposed.

The scientific objectives of Modal include the two following methodological directions: generative model design and data visualization through such models. In each case, several means of dissemination are considered towards academic and/or industrial communities: publications in international journals (in statistics or biostatistics), workshops to raise or identify emerging topics, and publicly available specific software relying on the proposed new methodologies.

3. Research Program

3.1. Generative model design

The first objective of MODAL consists in designing, analyzing, estimating and evaluating new generative parametric models for multivariate and/or heterogeneous data. It corresponds typically to continuous and categorical data but it includes also other widespread ones like ordinal, functional, ranks,...Designed models have to take into account potential correlations between variables while being (1) justifiable and realistic, (2) meaningful and parsimoniously parameterized, (3) of low computational complexity. The main purpose is to identify a few theoretical and general principles for model generation, loosely dependent on the variable nature. In this context, we propose two concurrent approaches which could be general enough for dealing with correlation between many types of homogeneous or heterogeneous variables:

- Designs general models by combining two extreme models (full dependent and full independent) which are well-defined for most of variables;
- Uses kernels as a general way for dealing with multivariate and heterogeneous variables.

3.2. Data visualization

The second objective of MODAL is to propose meaningful and quite accurate low dimensional visualizations of data typically in two-dimensional (2D) spaces, less frequently in one-dimensional (1D) or three-dimensional (3D) spaces, by using the generative models designed in the first objective. We propose also to visualize simultaneously the data and the model. All visualizations will depend on the aim at hand (typically clustering, classification or density estimation). The main originality of this objective lies in the use of models for visualization, a strategy from which we expect to have a better control on the subjectivity necessarily induced by any graphical display. In addition, the proposed approach has to be general enough to be independent on the variable nature. Note that the visualization objective is consistent with the dissemination of our methodologies through specific softwares. Indeed, displaying data is an important step in the data analysis process.

4. Application Domains

4.1. Multiple domains applications

Participants: Sophie Dabo, Cristian Preda, Vincent Vandewalle, Alain Celisse, Benjamin Guedj, Christophe Biernacki, Guillemette Marot.

Modal targets a wide spectrum of application domains.

In particular, several members are interested in classification of functional data and functional regression models when data are correlated (temporally or spatially) and application to hydrological, environnemental or medical data.

Other topics include any application domains involving clustering, prediction or visualization (such as image segmentation, (online) clustering in retail, failure prediction in the steel industry, sales prediction in retail, ...). In most cases, we enforce the use of probabilistic models with associated software.

4.2. Genomics

Participants: Guillemette Marot, Alain Celisse.

With the use of high throughput technologies, more and more data are generated in molecular biology studies. Our developments are applied at several levels:

- genomics to detect aberrations in genomic profiles from patients suffering from cancers
- transcriptomics to find differentially expressed genes, e.g. between ill and healthy patients
- epigenetics to better understand cells mechanisms

5. Highlights of the Year

5.1. Highlights of the Year

The major highlight of Modal is related to transfer of its research towards the private sector. In 2016, several major bilateral contracts have been signed between Modal and leading international companies based in Hauts-de-France. Those collaborations directly proceed from the fundamental research carried within the team (see Section "Bilateral Contracts and Grants with Industry").

6. New Software and Platforms

6.1. BlockCluster

Block Clustering
KEYWORDS: Statistic analysis - Clustering package
SCIENTIFIC DESCRIPTION

Simultaneous clustering of rows and columns, usually designated by biclustering, co-clustering or block clustering, is an important technique in two way data analysis. It consists of estimating a mixture model which takes into account the block clustering problem on both the individual and variables sets. The blockcluster package provides a bridge between the C++ core library and the R statistical computing environment. This package allows to co-cluster binary, contingency, continuous and categorical data-sets. It also provides utility functions to visualize the results. This package may be useful for various applications in fields of Data mining, Information retrieval, Biology, computer vision and many more.

FUNCTIONAL DESCRIPTION

BlockCluster is an R package for co-clustering of binary, contingency and continuous data based on mixture models.

- Participants: Parmeet Bhatia, Serge Iovleff, Vincent Brault, Christophe Biernacki, Gilles Celeux and Vincent Kubicki
- Partner: Université de Technologie de Compiègne
- Contact: Serge Iovleff
- URL: http://cran.r-project.org/web/packages/blockcluster/index.html

6.2. Clustericat

FUNCTIONAL DESCRIPTION

Clustericat is an R package for model-based clustering of categorical data. In this package, the Conditional Correlated Model (CCM), published in 2014, takes into account the main conditional dependencies between variables through extreme dependence situations (independence and deterministic dependence). Clustericat performs the model selection and provides the best model according to the BIC criterion and the maximum likelihood estimates.

- Participants: Matthieu Marbac-Lourdelle, Vincent Vandewalle and Christophe Biernacki
- Contact: Matthieu Marbac-Lourdelle
- URL: https://r-forge.r-project.org/R/?group_id=1803

6.3. CoModes

FUNCTIONAL DESCRIPTION

CoModes is another R package for model-based clustering of categorical data. In this package, the Conditional Modes Model (CMM) (published in 2016) takes into account the main conditional dependencies between variables through particular modality crossings (so-called modes). CoModes performs the model selection and provides the best model according to the exact integrated likelihood criterion and the maximum likelihood estimates.

- Participants: Matthieu Marbac-Lourdelle, Vincent Vandewalle and Christophe Biernacki
- Contact: Matthieu Marbac-Lourdelle
- URL: https://r-forge.r-project.org/R/?group_id=1809

6.4. CorReg

FUNCTIONAL DESCRIPTION

The main idea of the CorReg package is to consider some form of sub-regression models, some variables defining others. We can then remove temporarily some of the variables to overcome ill-conditioned matrices inherent in linear regression and then reinject the deleted information, based on the structure that links the variables. The final model therefore takes into account all the variables but without suffering from the consequences of correlations between variables or high dimension.

- Participants: Clément Thery and Christophe Biernacki
- Contact: Clément Thery
- URL: https://cran.r-project.org/web/packages/CorReg/index.html

6.5. FunFEM

FUNCTIONAL DESCRIPTION

FunFEM package for R proposes a clustering tool for functional data. The model-based algorithm clusters the functional data into discriminative subspaces.

- Participants: Charles Bouveyron and Julien Jacques
- Contact: Charles Bouveyron
- URL: https://cran.r-project.org/web/packages/funFEM/index.html

6.6. FunHDDC

FUNCTIONAL DESCRIPTION

FunHDDC package for R proposes a clustering tool for functional data. The model-based clustering algorithm considers that functional data live in cluster-specific subspaces.

- Participants: Charles Bouveyron and Julien Jacques
- Contact: Charles Bouveyron
- URL: https://cran.r-project.org/web/packages/funHDDC/index.html

6.7. Galaxy - MPAgenomics

KEYWORDS: Bioinformatics - Data mining - Statistics - Genomics
FUNCTIONAL DESCRIPTION

Galaxy is an open, web-based platform for data intensive biomedical research. Galaxy features user friendly interface, workflow management, sharing functionalities and is widely used in the biologist community. The MPAgenomics R package developped by MODAL has been integrated into Galaxy, and the Galaxy MODAL instance has been publicly deployed thanks to the IFB-cloud infrastructure.

- Participants: Guillemette Marot and Samuel Blanck
- Contact: Guillemette Marot
- URL: https://cloud.france-bioinformatique.fr/accounts/login/

6.8. HDPenReg

High-Dimensional Penalized Regression
FUNCTIONAL DESCRIPTION

HDPenReg is an R-package based on a C++ code dedicated to the estimation of regression model with l1-penalization.

- Participants: Quentin Grimonprez and Serge Iovleff
- Contact: Quentin Grimonprez
- URL: https://cran.r-project.org/web/packages/HDPenReg/index.html

6.9. MPAGenomics

Multi-Patient Analysis of Genomic markers
KEYWORDS: Segmentation - Genomics - Marker selection - Biostatistics
SCIENTIFIC DESCRIPTION

MPAgenomics is an R package for multi-patients analysis of genomics markers. It enables to study several copy number and SNP data profiles at the same time. It offers wrappers from commonly used packages to offer a pipeline for beginners in R. It also proposes a special way of choosing some crucial parameters to change some default values which were not adapted in the original packages. For multi-patients analysis, it wraps some penalized regression methods implemented in HDPenReg.
FUNCTIONAL DESCRIPTION

MPAgenomics provides functions to preprocess and analyze genomic data. It is devoted to: (i) efficient segmentation and (ii) genomic marker selection from multi-patient copy number and SNP data profiles.

- Participants: Quentin Grimonprez, Guillemette Marot and Samuel Blanck
- Contact: Guillemette Marot
- URL: https://cran.r-project.org/web/packages/MPAgenomics/index.html

6.10. MetaMA

Meta-analysis for MicroArrays
KEYWORDS: Transcriptomics - Meta-analysis - Differential analysis - Microarrays - Biostatistics
FUNCTIONAL DESCRIPTION

MetaMA is a specialised software for microarrays. It is an R package which combines either p-values or modified effect sizes from different studies to find differentially expressed genes. The main competitor of metaMA is geneMeta. Compared to geneMeta, metaMA offers an improvement for small sample size datasets since the corresponding modelling is based on shrinkage approaches.

- Participant: Guillemette Marot
- Contact: Guillemette Marot
- URL: https://cran.r-project.org/web/packages/metaMA/index.html

6.11. MetaRNASeq

KEYWORDS: Transcriptomics - Meta-analysis - Differential analysis - High throughput sequencing - Biostatistics
FUNCTIONAL DESCRIPTION

This is joint work with Andrea Rau (INRA, Jouy-en-Josas). MetaRNASeq is a specialised software for RNA-seq experiments. It is an R package which is an adaptation of the MetaMA package presented previously. Both implement the same kind of methods but specificities of the two types of technologies require some adaptations to each one.

- Participants: Guillemette Marot and Andrea Rau
- Contact: Guillemette Marot
- URL: https://cran.r-project.org/web/packages/metaRNASeq/index.html

6.12. MixAll

Clustering using Mixture Models
KEYWORDS: Clustering - Clustering package - Generative Models
FUNCTIONAL DESCRIPTION

MixAll is a model-based clustering package for modelling mixed data sets. It has been engineered around the idea of easy and quick integration of any kind of mixture models for any kind of data, under the conditional independence assumption. Currently five models (Gaussian mixtures, categorical mixtures, Poisson mixtures, Gamma mixtures and kernel mixtures) are implemented. MixAll has the ability to natively manage completely missing values when assumed as random. MixAll is used as an R package, but its internals are coded in C++ as part of the STK++ library (http://www.stkpp.org) for faster computation.

- Participant: Serge Iovleff
- Contact: Serge Iovleff
- URL: https://cran.r-project.org/web/packages/MixAll/

6.13. MixCluster

FUNCTIONAL DESCRIPTION

MixCluster is an R package for model-based clustering of mixed data (continuous, binary, integer). In this package, the model, submitted for publication in 2014, takes into account the main conditional dependencies between variables through Gaussian copula. Mixcluster performs the model selection and provides the best model according to Bayesian approaches.

- Participants: Matthieu Marbac-Lourdelle, Christophe Biernacki and Vincent Vandewalle
- Contact: Matthieu Marbac-Lourdelle
- URL: https://r-forge.r-project.org/R/?group_id=1939

6.14. Mixmod

Many-purpose software for data mining and statistical learning
KEYWORDS: Data mining - Classification - Mixed data - Data modeling - Big data
FUNCTIONAL DESCRIPTION

Mixmod is a free toolbox for data mining and statistical learning designed for large and high dimensional data sets. Mixmod provides reliable estimation algorithms and relevant model selection criteria.

It has been successfully applied to marketing, credit scoring, epidemiology, genomics and reliability among other domains. Its particularity is to propose a model-based approach leading to a lot of methods for classification and clustering.

Mixmod allows to assess the stability of the results with simple and thorough scores. It provides an easy-to-use graphical user interface (mixmodGUI) and functions for the R (Rmixmod) and Matlab (mixmodForMatlab) environments.

- Participants: Christophe Biernacki, Gilles Celeux, Gérard Govaert, Florent Langrognet, Serge Iovleff, Remi Lebret and Benjamin Auder
- Partners: CNRS - HEUDIASYC - Laboratoire Paul Painlevé - LIFL - LMB - Université Lille 1
- Contact: Gilles Celeux
- URL: http://www.mixmod.org

6.15. MixtComp

Mixture Computation
KEYWORDS: Clustering - Statistics - Missing data

FUNCTIONAL DESCRIPTION

MixtComp (Mixture Computation) is a model-based clustering package for mixed data originating from the Modal team (Inria Lille). It has been engineered around the idea of easy and quick integration of all new univariate models, under the conditional independence assumption. New models will eventually be available from researches, carried out by the Modal team or by other teams. Currently, central architecture of MixtComp is built and functionality has been field-tested through industry partnerships. Three basic models (Gaussian, multinomial, Poisson) are implemented, as well as two advanced models (Ordinal and Rank). A new advanced model concerning functional data is also available in 2016. MixtComp has the ability to natively manage missing data (completely or by interval). MixtComp is used as an R package, but its internals are coded in C++ using state of the art libraries for faster computation.

- Participants: Vincent Kubicki, Christophe Biernacki and Serge Iovleff
- Contact: Christophe Biernacki
- URL: https://massiccc.lille.inria.fr/#/

6.16. RankCluster

FUNCTIONAL DESCRIPTION

Rankcluster package for R proposes a clustering tool for ranking data. Multivariate and partial rankings can be also taken into account. Rankcluster now supports tied ranking data.

- Participants: Christophe Biernacki, Julien Jacques and Quentin Grimonprez
- Contact: Quentin Grimonprez
- URL: https://cran.r-project.org/web/packages/Rankcluster/index.html

6.17. STK++

Statistical ToolKit
KEYWORDS: Statistics - Linear algebra - Framework
FUNCTIONAL DESCRIPTION

STK++ (Statistical ToolKit in C++) is a versatile, fast, reliable and elegant collection of C++ classes for statistics, clustering, linear algebra, arrays (with an API Eigen-like), regression, dimension reduction, etc. The library is interfaced with lapack for many linear algebra usual methods. Some functionalities provided by the library are available in the R environment using rtkpp and rtkore.

STK++ is suitable for projects ranging from small one-off projects to complete data mining application suites.

- Participant: Serge Iovleff
- Contact: Serge Iovleff
- URL: http://www.stkpp.org

6.18. clere

FUNCTIONAL DESCRIPTION

The clere package for R proposes variable clustering in high dimensional linear regression. Available on CRAN and now published to an international journal dedicated to software: [24].

- Participants: Loïc Yengo, Christophe Biernacki and Julien Jacques
- Contact: Loïc Yengo
- URL: https://cran.r-project.org/web/packages/clere/index.html

6.19. rtkore

STK++ core library integration to R using Rcpp

KEYWORDS: C++ - Data mining - Clustering - Statistics - Regression
FUNCTIONAL DESCRIPTION

STK++ (http://www.stkpp.org) is a collection of C++ classes for statistics, clustering, linear algebra, arrays (with an Eigen-like API), regression, dimension reduction, etc. The integration of the library to R is using Rcpp. The rtkore package includes the header files from the STK++ core library. All files contain only templated classes or inlined functions. STK++ is licensed under the GNU LGPL version 2 or later. rtkore (the stkpp integration into R) is licensed under the GNU GPL version 2 or later. See file LICENSE.note for details.

- Participant: Serge Iovleff
- Contact: Serge Iovleff
- URL: https://cran.r-project.org/web/packages/rtkore/index.html

7. New Results

7.1. An oracle inequality for Quasi-Bayesian Non-Negative Matrix Factorisation

Participant: Benjamin Guedj.

We have extended the quasi-Bayesian perspective to the popular setting of non-negative matrix factorisation. This is a pivotal problem in machine learning (image segmentation, recommendation systems, audio source separation, ...) and we were able to propose an original estimator of the unobserved matrix. An oracle inequality is derived, along with several possible implementations. This work is now submitted to an international journal [38].

Joint work with Pierre Alquier.

7.2. PAC-Bayesian Online Clustering

Participants: Benjamin Guedj, Le Li.

We have extended the PAC-Bayesian framework to online learning. Our algorithm (called PACBO) performs online clustering of random sequences, and is supported by strong theoretical (regret bounds) and algorithmic (ergodicity of an MCMC implementation) results. This work is now submitted to an international journal [46].

Joint work with Sébastien Loustau.

7.3. Simpler PAC-Bayesian Bounds for Hostile Data

Participant: Benjamin Guedj.

We have introduced an original and much simpler way of deriving PAC-Bayesian bounds, through the use of f-divergences (therefore generalizing earlier works on Renyi's divergence and Kullback-Leibler divergence). This work is now submitted to an international conference [39].

Joint work with Pierre Alquier.

7.4. Clustering categorical functional data: Application to medical discharge letters

Participants: Cristian Preda, Cristina Preda, Vincent Vandewalle.

Categorical functional data represented by paths of a stochastic jump process are considered for clustering. For paths of the same length, the extension of the multiple correspondence analysis allows the use of well-known methods for clustering finite dimensional data. When the paths are of different lengths, the analysis is more complex. In this case, for Markov models we have proposed an EM algorithm to estimate a mixture of Markov processes. This work has been presented in a workshop [48].

7.5. Simultaneous dimension reduction and multi-objective clustering

Participant: Vincent Vandewalle.

In model based clustering of quantitative data it is often supposed that only one clustering variable explains the heterogeneity of all the others variables. However, when variables come from different sources, it is often unrealistic to suppose that the heterogeneity of the data can only be explained by one variable. If such an assumption is made, this could lead to a high number of clusters which could be difficult to interpret. A model based multi-objective clustering is proposed, it assumes the existence of several latent clustering variables, each one explaining the heterogeneity of the data on some clustering projection. In order to estimate the parameters of the model an EM algorithm is proposed, it mainly relies on a reinterpretation of the standard factorial discriminant analysis in a probabilistic way. The obtained results are projections of the data on some principal clustering components allowing some synthetic interpretation of the principal clusters raised by the data. This work has been presented in a conference [49].

7.6. Spatial Prediction of solar energy

Participant: Sophie Dabo.

Sophie Dabo-Niang's new result concern a work on spatial prediction of solar Energy in collaboration with some physicians and is now published [15].

This paper introduces a new approach for the forecasting of solar radiation series at a located station for very short time scale. We built a multivariate model in using few stations (3 stations). The proposed model is a spatio temporal vector autoregressive VAR model specifically designed for the analysis of spatially sparse spatio-temporal data. This model differs from classic linear models in using spatial and temporal parameters where the available predictors are the lagged values at each station. A spatial structure of stations is defined by the sequential introduction of predictors in the model. Moreover, an iterative strategy in the process of our model will select the necessary stations removing the uninteresting predictors and also selecting the optimal p-order. We studied the performance of this model. The metric error, the relative root mean squared error (rRMSE), is presented at different short time scales. Moreover, we compared the results of our model to simple and well known persistence model and those found in literature.

7.7. Multiple change-point detection

Participants: Alain Celisse, Guillemette Marot.

This is a joint work with Morgane Pierre-Jean and Guillem Rigaill (Univ. Evry).

The paper related to the work described in previous MODAL team reports (sections Kernel change point) has been pursuied and made available on Arxiv [42]. For recall, this work focuses on the problem of detecting abrupt changes arising in the full distribution of the observations (not only in the mean or variance). It provides greatly improved algorithms in terms of computational complexity (both in time and space). The computational and statistical performances of these new algorithms have been assessed through empirical experiments, which are detailed in the preprint.

7.8. Differential gene expression analysis

Participants: Alain Celisse, Guillemette Marot.

The use of empirical Bayesian techniques implemented in the R package metaMA has enabled to better understand Waldenstrom's macroglobulinemia. The new findings in Biology have been published in [18].

7.9. New concentration inequalities for the leave-p-out CV estimator

Participant: Alain Celisse.

New concentration inequalities have been established for the leave-p-out cross-validation estimator applied to assess the performance the k-nearest neighbour binary classifier. Joint work with Tristan Mary-Huard.

7.10. A new notion of stability for learning algorithms

Participants: Alain Celisse, Benjamin Guedj.

We introduced a new notion of stability for learning algorithms, which bridges the gap between the earlier uniform and hypothesis stability notions. It allows us to derive new PAC exponential concentration inequalities that apply to the Ridge regression algorithm as a first step. The first version of this work is presented in the preprint [41] and is now an active line of research.

7.11. Model for conditionally correlated categorical data

Participants: Christophe Biernacki, Matthieu Marbac Lourdelle, Vincent Vandewalle.

It is a model-based clustering proposal (called CMM for Conditional Modes Model) where categorical data are grouped into conditionally independent blocks. The corresponding block distribution is a parsimonious multinomial distribution where the few free parameters correspond to the most likely modality crossings, while the remaining probability mass is uniformly spread over the other modality crossings. The exact computation of the integrated complete-data likelihood allows to perform the model selection, by a Gibbs sampler, reducing the computing time consuming by parameter estimation and avoiding BIC criterion biases pointed out by our experiments. This work is now published in the international journal Advances in Data Analysis and Classification (Marbac et al, 2016). Furthermore, an R package (CoModes) is available on Rforge.

7.12. Mixture model for mixed kind of data

Participants: Christophe Biernacki, Matthieu Marbac Lourdelle, Vincent Vandewalle.

A mixture model of Gaussian copula allows to cluster mixed kind of data. Each component is composed by classical margins while the conditional dependencies between the variables is modeled by a Gaussian copula. The parameter estimation is performed by a Gibbs sampler. This work has been now accepted to an international journal [21]. Furthermore, an R package (MixCluster) is available on Rforge.

7.13. Degeneracy in multivariate Gaussian mixtures (complete data case)

Participant: Christophe Biernacki.

In the case of Gaussian mixtures, unbounded likelihood is an important theoretical and practical problem. Using the weak information that the latent sample size of each component has to be greater than the space dimension, a simple non-asymptotic stochastic lower bound on variances is derived. It is proved also that maximizing the likelihood under this data-driven constraint leads to consistent estimates. This work has been presented as an invited talk to the international workshop [28] and a paper for an international journal is been prepared.

This is a joined work with Gwënaelle Castellan of University of Lille.

7.14. Degeneracy in multivariate Gaussian mixtures (missing data case)

Participants: Christophe Biernacki, Vincent Vandewalle.

In the case of multivariate Gaussian mixtures, unbounded likelihood is an important theoretical and practical problem. However, in the case of missing data situations, this drawback is exacerbated for too reasons. Firstly, degeneracy frequence increases with missing data occurrence. Secondly, the EM dynamic is hardly detected since it implies linear grows of the log-likelihood, contrary to exponential grows in the complete data case, leading to computation waste and also high risk of erroneous estimates. Using the weak information that the latent sample size of each component (restricted to complete data) has to be greater than the space dimension, it is derived a simple contraint EM algorithm variant allowing to solve simultaneously both problems. This work has been presented to the international workshop [28] and a paper for an international journal is been prepared.

7.15. Data units selection in statistics

Participant: Christophe Biernacki.

Usually, the data unit definition is fixed by the practitioner but it can happen that it hesitates between several data unit options. In this context, it is highlighted that it is possible to embed data unit selection into a classical model selection principle. The problem is introduced in a regression context before to focus on the model-based clustering and co-clustering context, for data of different kinds (continuous, categorical, counting, ...). It has led to an invitation to an international workshop [29] and a preprint is being to be prepared.

It is a joint work with Alexandre Lourme from University of Bordeaux.

7.16. Label switching in Bayesian mixture model estimation

Participants: Christophe Biernacki, Benjamin Guedj, Vincent Vandewalle.

In the case of mixtures of distributions, it is well-known that the Bayesian posterior distribution is invariant to label switching, it means invariant to any renumbering of components. Consequences are important, typically leading to unuseful estimates like the posterior mean. Many attempts exist to solve this problem but it is advocated in this work that such a quest should be unfruitful since it is a direct consequence of the label non-identifiability of mixtures themselves. The present work proposes an original way to manage the label switching problem based on the Gibbs algorithm dynamic. The basic idea is to control the label switching probability along Gibbs iterations, controlled by both the sample size and the component overlap. An early version of this work has been presented as an invited talk to the international workshop [28].

7.17. Trade-off computation time and accuracy

Participants: Christophe Biernacki, Maxime Brunin, Alain Celisse.

Most estimates practically arise from algorithmic processes aiming at optimizing some standard, but usually only asymptotically relevant, criteria. Thus, the quality of the resulting estimate is a function of both the iteration number and also the involved sample size. An important question is to design accurate estimates while saving computation time, and we address it in the simplified context of linear regression here. Fixing the sample size, we focus on estimating an early stopping time of a gradient descent estimation process aiming at maximizing the likelihood. It appears that the accuracy gain of such a stopping time increases with the number of covariates, indicating potential interest of the method in real situations involving many covariates. A first version of this work has been presented to an international conference [27], and a preprint is being in progress.

7.18. Projection under pairwise control

Participant: Christophe Biernacki.

Visualization of high-dimensional and possibly complex (non continuous for instance) data onto a low-dimensional space may be difficult. Several projection methods have been already proposed for displaying such high-dimensional structures on a lower-dimensional space, but the information lost is not always easy to use. Here, a new projection paradigm is presented to describe a non-linear projection method that takes into account the projection quality of each projected point in the reduced space, this quality being directly available in the same scale as this reduced space. More specifically, this novel method allows a straightforward visualization data in R^2 with a simple reading of the approximation quality, and provides then a novel variant of dimensionality reduction.

This work is under revision in an international journal [37] and it has also been presented to an international conference [25].

It is a joint work with Hiba Alawieh and Nicolas Wicker, both from University of Lille.

7.19. Matching of descriptors evolving over time

Participants: Christophe Biernacki, Anne-Lise Bedenel.

In the web domain, and in particular for insurance comparison, data constantly evolve, implying that it is difficult to directly exploit them. For example, to do a classification, performing standard learning processes require data descriptor equal for both learning and test samples. Indeed, for answering to web surfer expectation, online forms whence data come from are regularly modified. So, features and data descriptors are also regularly modified. In this work, it is introduced a process to estimate and understand connections between transformed data descriptors. This estimated matching between descriptors will be a preliminary step before applying later classical learning methods. This work has been presented to a national conference [33], with international audience.

It is a joint work with Laetitia Jourdan, from University of Lille and Inria.

7.20. Real-time audio sources classification

Participants: Christophe Biernacki, Maxime Baelde.

Recent research on machine learning focuses on audio source identification in complex environments. They rely on extracting features from audio signals and use machine learning techniques to model the sound classes. However, such techniques are often not optimized for a real-time implementation and in multi-source conditions. It is proposed here a new real-time audio single-source classification method based on a dictionary of sound models (that can be extended to a multi-source setting). The sound spectrums are modeled with mixture models and form a dictionary. The classification is based on a comparison with all the elements of the dictionary by computing likelihoods and the best match is used as a result. It is found that this technique outperforms classic methods within a temporal horizon of 0.5s per decision (achieved 6% of errors on a database composed of 50 classes). Future works will focus on the multi-sources classification and reduce the computational load. This work has been accepted in 2016 to be presented in 2017 to an international conference in Signal Processing [32].

It is a joint work with Raphaël Greff, from the A-Volute company.

7.21. Model-Based Co-clustering for Ordinal Data

Participants: Christophe Biernacki, Julien Jacques.

A model-based coclustering algorithm for ordinal data is presented. This algorithm relies on the latent block model using the BOS model (Biernacki and Jacques, 2015, Stat. Comput.) for ordinal data and a SEM-Gibbs algorithm for inference. Numerical experiments on simulated data illustrate the efficiency of the inference strategy. This work has been presented to an international workshop [30] and also to a national conference with an international audience [35].

7.22. Computational and statistical trade-offs in change-point detection

Participants: Christophe Biernacki, Maxime Brunin, Alain Celisse.

The change-point detection problem aims to detect changes in the distribution of observations collected over the time between the instants 1,...,T in the offline context. These changes occur at some instants called change-points. Our method provides consistent estimates of the change-points obtained by the Kernel Binary Segmentation algorithm with stopping rule (KBS). Moreover, the proposed method has a lower complexity in time and in space than the Kernel Dynamic Programming (KDP). This work has been presented to a national conference with an international audience [34].

7.23. MixtComp software for full mixed data

Participants: Christophe Biernacki, Vincent Kubicki.

MixtComp (Mixture Computation) is an integration software from the MODAL team for model-based clustering of mixed data. Its computing core is written in C++ and is accessed through an R interface. Its architecture allows to easily and quickly integrate new univariate models (under the conditional independence assumption) as they are published. The first phase of development was the implementation of three basic models (Gaussian, Multinomial, Poisson) with the native management of partially observed data (including intervals). It now implements models related to ordinal data (2015), rank data (2015) and functional data (2016), still with missing or partially missing data. The code is developed internally, and has been field-tested through several contracted partnerships (see the section about contracts). It is now referenced in the BIL database and the APP. It is available through a new web interface, called MASSICCC at https://massiccc.lille.inria.fr/#/ (see also the dedicated section). MixtComp has been presented to an invited talk in October 2016 at the Academy of Sciences in Tunisia [26].

7.24. MASSICCC platform for SaaS software availability

Participants: Christophe Biernacki, Vincent Kubicki, Matthieu Marbac Lourdelle.

MASSICCC is a demonstration platform giving access through a SaaS (service as a software) concept to data analysis libraries developed at Inria. It allows to obtain results either directly through a website specific display (specific and interactive visual outputs) or through an R data object download. It started in October 2015 for two years and is common to the Modal team (Inria Lille) and the Select team (Inria Saclay). In 2016, two packages have been integrated: Mixmod and MixtComp (see the specific section about MixtComp). In 2017, it is planned to integrate the BlockCluster package. The MASSICCC platform gardually replaces the former BigStat platform available here: https://modal-research.lille.inria.fr/BigStat/. BigStat and MASSICCC have been both presented to an invited talk in October 2016 at the Academy of Sciences in Tunisia [26].

MASSICCC has led to a first short meeting in April 2016 in Lille for obtaining a feedback from company and academic users. Here is the link towards this event: Link. A second similar event is planned in February 2017 in Paris. Joint work with Jonas Renault and Josselin Demont (both at InriaTech).

The MASSICCC platform is available on https://massiccc.lille.inria.fr

7.25. CoModes package for correlated categorical variables

Participants: Christophe Biernacki, Matthieu Marbac Lourdelle, Vincent Vandewalle.

CoModes is an R package for model-based clustering of categorical data. In this package, the Conditional Modes Model (CMM), published in 2016 (Marbac et al, 2016), takes into account the main conditional dependencies between variables through particular modality crossings (so-called modes). CoModes performs the model selection and provides the best model according to the exact integrated likelihood criterion and the maximum likelihood estimates. It is available online on Rforge (https://r-forge.r-project.org/R/?group_id=1809).

7.26. MixCluster package for correlated mixed variables

Participants: Christophe Biernacki, Matthieu Marbac Lourdelle, Vincent Vandewalle.

MixCluster is an R package for model-based clustering of mixed data (continuous, binary, integer). In this package, the model, accepted for publication in 2016 [21], takes into account the main conditional dependencies between variables through Gaussian copula. Mixcluster performs the model selection and provides the best model according to Bayesian approaches. It is available online on Rforge (https://r-forge.r-project.org/R/?group_id=1939).

8. Bilateral Contracts and Grants with Industry

8.1. Arcelor-Mittal

Participant: Christophe Biernacki.

Arcelor-Mittal is a leader company in steel industry. This 11 months contract aims at optimizing predictive maintenance from mixed data (continuous, categorical, functional) provided by multiple sensors disseminated in steel production lines.

It is a joint work with Martin Bue and Vincent Kubicki (InriaTech engineers).

8.2. Banque Accord

Participants: Christophe Biernacki, Vincent Vandewalle.

Banque Accord is a credit scoring company. This 3 months contract aims at improving credit scoring performance by using the clustering principle inside the predictive process. In addition, directly managing mixed data (continuous, categorical, missing) has to be taken into account.

It is a joint work with Quentin Grimonprez (InriaTech engineer).

8.3. Vallourec

Participant: Christophe Biernacki.

Vallourec is a world leader in premium tubular solutions for the energy markets and for other demanding industrial applications. This 9 months contract aims at predicting quality of tubular connections from mixed data (continuous, categorical, functional).

It is a joint work with Vincent Kubicki (InriaTech engineer).

8.4. Cylande

Participants: Christophe Biernacki, Vincent Vandewalle.

Cylande is a software editor for retail. This 12 months contract aims at predicting future sales from past sales, including also many other available information.

It is a joint work with Etienne Goffinet and Vincent Kubicki (InriaTech engineers).

8.5. NFID

Participants: Benjamin Guedj, Quentin Grimonprez.

NFID is the agency dedicated to innovation policies of the Hauts-de-France region.

This 3 months contract aims at clustering companies from Hauts-de-France based on their economic, social, environmental, innovation, activities data. The proposed methodology relies on the MixtComp software developed within Modal, and allows for the creation of a predictive analysis tool for NFID. This predictive tool aims at identifying regional companies with the highest innovative abilities, and has a great economic and politic impact.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. L'impact de l'évolution de l'état émotionnel et cognitif ressenti sur la reprise de l'activité de femmes atteintes d'un cancer du sein (Protocole FACEBROK)

Participant: Sophie Dabo.

Partners: LAPCOS (EA 7278), UMR 9193 SCALab, LEM UMR 9221 LEM, Modal-Inria, EA CRDP

9.1.2. Main partners of bilille

Participant: Guillemette Marot.

bilille, the bioinformatics platform of Lille officially gathers from Nov. 2015 a few bioinformaticians, biostatisticians and bioanalysts from the following teams:

EA2694 (Univ. Lille 2, CHRU, Inria)

FRABIO, FR3688 (Univ. Lille 1,CNRS)

CBP / GFS (Univ. Lille 2, CHRU)

TAG (Univ. Lille 2, CNRS, INSERM, Institut Pasteur de Lille)

U1167 (Univ. Lille 2, CHRU, INSERM et Institut Pasteur de Lille)

U1011 (Univ. Lille 2, INSERM)

UMR8198 (Univ. Lille 1, CNRS)

LIGAN PM (Univ. Lille 2, CNRS)

BONSAI (Inria, Univ. Lille 1, CNRS)

Those teams are thus the main partners of MODAL concerning biostatistics for bioinformatics. Guillemette Marot is the leader of the platform and works in close collaboration with the following people for the leadership of the scientific strategy related to the platform:

H. Touzet, BONSAI (deputy head of bilille)

P. Touzet, UMR8198 (deputy head of bilille)

V. Chouraki, U1167

M. Figeac, CBP / GFS

D. Hot, TAG

V. Leclère, Insitut Charles Viollette

M. Lensink, UGSF

9.1.3. New collaborations of the year linked to bilille, the bioinformatics and bioanalysis platform

Participants: Guillemette Marot, Samuel Blanck.

Guillemette Marot has supervised the data analysis part or support in biostatistics tools testing for the following research projects involving Samuel Blanck or engineers from bilille (only the names of the principal investigators of the project are given even if several partners are sometimes involved in the project):

U 1011, H. Duez, circadiomics project

CIIL, J.C. Sirard, Flagnew project

JPARC, M.H. David, biostatistics related to DNAse-seq

9.1.4. Collaboration linked to SIRIC Oncolille

Participants: Sophie Dabo, Guillemette Marot.

During the 'Plan Cancer 2' period, eight SIRICs ('Site de Recherche Intégrée sur le Cancer') were created in France, including the SIRIC ONCOLille (Link). More recently, the SFR Cancer has been created and Sophie Dabo-Niang is a member of the board that aims to create an Interdisciplinary Cancer Research Institute in Lille, based on ONCOLille. Guillemette Marot is still involved in several collaborations linked to cancer, through the projects analysed by the bilille platform.

9.2. National Initiatives

9.2.1. Programme of Investments for the Future (PIA)

Bilille is a member of two PIA "Infrastructures en biologie-santé":

France Génomique https://www.france-genomique.org/spip/?lang=en

IFB (French Institute of Bioinformatics) https://www.france-bioinformatique.fr/en

As leader of the platform, Guillemette Marot is thus involved in these networks.

9.2.2. Working groups

Sophie Dabo-Niang belongs to the following working groups.

- STAFAV (STatistiques pour l'Afrique Francophone et Applications au Vivant)
- ERCIM Working Group on computational and Methodological Statistics, Nonparametric Statistics Team
- Ameriska

Benjamin Guedj belongs to the following working groups (GdR) of CNRS: ISIS (local referee for Inria Lille - Nord Europe), MaDICS, MASCOT-NUM (local referee for Inria Lille - Nord Europe).

Guillemette Marot belongs to the StatOmique working group.

9.2.3. ANR

Participant: Cristian Preda.

ClinMine Project-2014-2017

ANR project (ANR TECSAN - Technologie de la santé)

Main coordinator of the project: Clarisse Dhaenens, CRIStAL, USTL

7 partners - EA 1046 (Maladie d'Alzheimer et pathologies vasculaires, Faculté de Médecine, Lille), EA 2694 (Centre d'Etudes et de Recherche en Informatique Médicale - Faculté de Médecine, Lille), MODAL (Inria LNE), Alicante (Entreprise), CHRU de Montpelier, GHICL (Groupe Hospitalier de l'Institut Catholique de Lille), CRIStAL, USTL

9.2.4. Other initiatives

Serge Iovleff is the head of the project CloHe granted in 2016 by the Mastodons CNRS challenge "Big data and data quality". The project is axed on the design of classification and clustering algorithms for mixed data with missing values with applications to high spatial resolution multispectral satellite image time-series. Website.

9.3. European Initiatives

9.3.1. European Research Council

Benjamin Guedj has participated in the 2017 Starting call of the European Research Council (ERC), by submitting a project (called BEAGLE, standing for PAC-Bayesian Agnostic Learning) in October 2016.

9.3.2. Collaborations with Major European Organizations

EMS (European Mathematical Society), Sophie Dabo-Niang

Nominated (November 2016) as member of EMS-CDC (Committee of Developing counties)

CIMPA (International Center of Pure and Applied Mathematics), Sophie Dabo-Niang

Nominated (June 2016) as member

9.4. International Initiatives

9.4.1. Inria International Labs

Sophie Dabo-Niang is a member of SIMERGE, a LIRIMA project-team started in January 2015. It includes researchers from Mistis (Inria Grenoble - Rhône-Alpes, France), LERSTAD (Laboratoire d'Etudes et de Recherches en Statistiques et Développement, Université Gaston Berger, Sénégal), IRD (Institut de Recherche pour le Développement, Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Dakar, Sénégal) and LEM lab (Lille Economie et Management, Université Lille 1, 2, 3, Modal, Inria Lille Nord-Europe, France).

9.4.2. Inria Associate Teams Not Involved in an Inria International Labs

Benjamin Guedj and Christophe Biernacki began a two years collaboration as "Equipes associées nord-européennes" with the Irish team "INSIGHT". The Centre for Data Analytics INSIGHT is about the size of Inria Lille - Nord Europe and is the main Irish research facility in Statistics and Machine Learning. It is focused on the next generation of machine learning (ML) and statistics (Stat) algorithms that can operate on large-scale, dynamic data. Nial Friel is the leader of the ML/Stat axis of INSIGHT, Brendan Murphy is a professor. The topic of this project is to manage statistical models inflation by the means of model clustering.

9.4.2.1. Informal International Partners

Benjamin Guedj regularly collaborates with Olivier Wintenberger from Københavns Universitet (KU, Denmark).

Benjamin Guedj regularly collaborates with Sylvain Robbiano from University College London (UCL, UK).

9.5. International Research Visitors

9.5.1. Visits of International Scientists

9.5.1.1. Internships

Rohit Uttam Bhagwat

Date: June 2016 - July 2016

Institution: Indian Institute of Science Education and Research, Kolkata (India)

Supervisor: Vincent Vandewalle

Siddharth Sharma Siddharth

Date: Nov 2015 - May 2016

Institution: LNM Institute of Information Technology (India)

Supervisor: Guillemette Marot

Miguel Assuncao

Date: September 2016 - February 2017

Institution: University of Lille

Supervisor: Christophe Biernacki and Vincent Kubicki

Ghazouani Yannis

Date: Oct 2015 - Sept 2016
Institution: École Centrale Lille - VEKIA
Supervisor: Alain Celisse

Hamza Tajmouati
Date: Oct 2015 - Sept 2016
Institution: École Centrale Lille
Supervisor: Alain Celisse

Astha Gupta
Date: May 2016 - Jul 2016
Institution: BITS Pilani (India)
Supervisor: Benjamin Guedj

Bhargav Srinivasa Desikan
Date: Aug 2016 - Jul 2017
Institution: BITS Pilani (India)
Supervisor: Benjamin Guedj

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

Sophie Dabo-Niang has visited AIMS-Senegal (African Institute of Mathematical Sciences) and SIMERGE (Inria International Lab of University Gaston-Berger, Senegal) from July to mid-August, 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Christophe Biernacki co-organised a one-day meeting "Introduction aux modèles statistiques scalables: Modélisation, nouveaux paradigmes, ecosystème Big Data" on December 5th 2016 at IHP (Paris). The website is here https://bigdata-stat.sciencesconf.org/ and it brought together about 40 registered people (maximum number for this session).

Christophe Biernacki, Benjamin Guedj and Sophie Dabo-Niang were in the organizing committee of the workshop "Big Data: Modelling, Estimation and Selection" at École Centrale de Lille on June 9th and 10th 2016 (https://indico.math.cnrs.fr/event/830/).

Alain Celisse and Guillemette Marot co-organized SMPGD in Lille: The Statistical Methods for Post Genomic Data workshop is an annual meeting dedicated to statistical methods for post genomic data analysis. The aim of the workshop is to present works from mathematical to applied Statistics, but also new areas in high throughput Biology that could need new statistical developments. The workshop is usually organized around 3 to 4 invited speakers, and 3 to 4 invited sessions, and one session of contributed abstracts (oral presentations and posters). As SMPGD is especially interested in the statistical, mathematical, algorithmics or modelling questions raised by modern biology, presentations are expected to focus on these points.

Benjamin Guedj co-founded and co-organised a one-day workshop called Young Statisticians and Probabilists (YSP), in Paris in January 2016. The topics were sequential learning, random trees, random maps and random matrices theory. Nearly 80 PhD students, postdocs and young researchers attended.

Benjamin Guedj is the organizer of the Modal team scientific seminar.

10.1.1.2. Member of the Organizing Committees

Benjamin Guedj has been a member of the steering committee for the FEM (Forum Emploi Maths) in Paris in December 2016. The FEM is the largest mathematics jobs fair in France and gathers universities, students, graduates, companies and institutions. Over 2,000 people attended this edition.

Sophie Dabo-Niang has been a member of the steering committees for the following events.

- CIMPA Research School: "Statistical methods for evaluation of extreme risks": April, 5-15, 2016, St-Louis, Senegal
- Worksop: "Financial and actuarial Mathematics": July 11-15, 2016, AIMS-Mbour, Senegal
- The first AWMA (African Women in Mathematics Association) regional Forum, July 8-9, 2016, AIMS-Mbour, Senegal
- Session "EO075: Quantile regression models for dependent data ", "The 9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2016)" (Link), December 9-11, 2016, Seville, Spain.
- Session "Asymptotic properties in nonparametric spatial problems ", "Third conference of the International Society for NonParametric Statistics (ISNPS)", June, 11-16, 2016, Avignon, France.
- Workshop "Statistical methods for recurrent data ", November 7th, 2016, Link.
- Workshop, "Learning with functional data", October, 7th, 2016, Lille, France. Link

Cristian Preda has co-organized the International Workshop on Applied Probability 2016 (Link).

Vincent Vandewalle is a member of the animation team of the bilille platform (https://wikis.univ-lille1.fr/bilille/animation). He has co-organized two scientific days, one in June 2016 on metagenomic analysis and another one in November 2016 on systems biology.

Serge Iovleff, Cristian Preda and Vincent Vandewalle have organised a one day workshop at Lille in October 2016 about learning with functional data. During this workshop, a large scope of methods for learning with functional data with application to various domains has been presented (https://functional-data.univ-lille1.fr).

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Sophie Dabo-Niang has been a member of the scientific committees of two conferences:

- STAHY 2016 Workshop, September 26-27, Quebec, Canada.
- International colloquium on financial Econometrics, November, 18 -19, 2016, Rabat, Maroc. Link

10.1.2.2. Reviewer

Alain Celisse has acted as a reviewer for AISTATS 2016.

Benjamin Guedj has acted as a reviewer for NIPS 2016 and AISTATS 2017.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Cristian Preda is a member of the editorial boards of

- Methodology and Computing in Applied Probability (Associate Editor)
- Romanian Journal of Mathematics and Computer Science" (Associate Editor)

Christophe Biernacki is an Associate Editor of the North-Western European Journal of Mathematics (NWEJM).

Sophie Dabo-Niang is a member of the editorial board of *Revista Colombiana de Estadística*.

10.1.3.2. Reviewer - Reviewing Activities

The Modal team is involved in 34 scientific outlets, among which are the most prestigious international journals related to statistics.

1. Advances in Data Analysis and Classification (Christophe Biernacki)
2. Annals of Applied Statistics (Sophie Dabo-Niang)
3. Annals of Statistics (Alain Celisse)
4. Bernoulli (Alain Celisse, Cristian Preda)
5. BMC Research Notes (Christophe Biernacki)
6. BMC Medical Research Methodology (Benjamin Guedj)
7. Canadian Journal of Statistics (Christophe Biernacki)
8. Chemometrics (Cristian Preda)
9. Computational Statistics and Data Analysis (Christophe Biernacki, Sophie Dabo-Niang, Cristian Preda, Vincent Vandewalle)
10. Data Mining and Knowledge Discovery (Christophe Biernacki)
11. Electronic Journal of Statistics (Alain Celisse, Sophie Dabo-Niang)
12. ESAIM: Probability and Statistics (Sophie Dabo-Niang)
13. Journal de la SFdS (Christophe Biernacki)
14. Journal of Classification (Christophe Biernacki)
15. Journal of Computational and Graphical Statistics (Vincent Vandewalle)
16. Journal of Machine Learning Research (Christophe Biernacki, Alain Celisse)
17. Journal of Multivariate Analysis (Sophie Dabo-Niang, Benjamin Guedj)
18. Journal of Nonparametric statistics (Sophie Dabo-Niang)
19. Journal of Statistical Planning and Inference (Christophe Biernacki)
20. Journal of Statistical Software (Christophe Biernacki)
21. Journal of the American Statistical Association (Benjamin Guedj)
22. Journal of the Royal Statistical Society, series A (Benjamin Guedj)
23. Knowledge and Information Systems (Christophe Biernacki)
24. Mathematical Reviews (Benjamin Guedj)
25. Methodology and Computing in Applied Probability (Cristian Preda)
26. Metrika (Sophie Dabo-Niang)
27. Molecular Ecology Resources (Benjamin Guedj)
28. Neurocomputing (Benjamin Guedj)
29. Statistical Inference for Stochastic Processes (Sophie Dabo-Niang)
30. Statistical Methods and Applications (Sophie Dabo-Niang)
31. Statistics (Sophie Dabo-Niang)
32. Statistics and Computing (Serge Iovleff)
33. Statistics and Probability Letters (Sophie Dabo-Niang, Benjamin Guedj)
34. The American Statistician (Christophe Biernacki)

10.1.4. Invited Talks

Christophe Biernacki's talks in 2016:

- Working Group on Model-Based Clustering Summer Session, Paris, July 17-23, 2016, https://maths.ucd.ie/~brendan/wgmbc2016.html, [28]
- Workshop on Model-based Clustering and Classification, September 5-7, 2016, Catania (Italy), http://mbc2.unict.it/, [29]
- Académie des Sciences, des Lettres et des Arts, Journée Scientifique "Big Data & Data Science", October 28th 2016, Tunis (Tunisia), http://www.beitalhikma.tn/p7536/, [26]
- 9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2016, ERCIM 2016), 9-11 December 2016, University of Seville, Spain http://cmstatistics.org/CMStatistics2016/, [27]
- Talk to the seminar of the "Laboratoire de Mathématiques de Besançon", February 29th 2016

Sophie Dabo-Niang's talks in 2016:

- Environmental Econometrics Day, Spatial Risk estimation and Application to environmental data, April, 24, 2016, Rabat, Marocco.
- CRoNoS FDA, satellite workshop of Compstat2016 (Link), Functional Binary Choice Models With Choice-Based Sampling, August, 26-28, Oviedo, Spain.
- AAS/AMU symposium on "Current Research Trends in Mathematical Sciences and applications", May, 17-20, 2016, UNESCO Chair of Mathematics, National Mathematical Centre (NMC), Abuja Nigeria.
- Learning with functional data, Functional Binary Choice Models With Choice-Based Sampling, October, 7, 2016, Lille, France.

Benjamin Guedj has been invited to deliver a talk to the 48èmes Journées de Statistique (JdS) of the French Statistical Society (June 2016, Montpellier, France).

Serge Iovleff gave a lightning talk entitled "MixAll: Un logiciel de classification non-supervisée" to the 5ème Rencontres R 2016.

Cristian Preda's talks in 2016:

1. International Workshop on Applied Probability (IWAP2016), 20-23 June, 2016, Toronto, Canada
2. 48e Journées de Statistique de la Société Francaise de Statistique, Montpellier, June 2016
3. 147th ICB Seminar Tenth International Seminar on statistics and clinical practice, May 15 - 18, 2016, Warsaw, Poland
4. 19-th Conference of the Romanian Society of Statistics and Probability, Universitatea Tehnica de Constructii Bucuresti, 27 mai 2016

10.1.5. Leadership within the Scientific Community

Christophe Biernacki is the president (since 2012) of the data mining and learning group of the French statistical association (SFdS, http://www.sfds.asso.fr/).

Benjamin Guedj is the president (since 2016) of the Young Statisticians group of the French Statistical Society.

Benjamin Guedj has joined the board of AMIES, the French Agency fostering collaborations between mathematicians and the private sector.

Guillemette Marot is responsible of bilille, the bioinformatics and bioanalysis platform of Lille. More information about the platform is available at https://wikis.univ-lille1.fr/bilille/

10.1.6. Scientific Expertise

Christophe Biernacki acted as an expert for two HCERES committees: one for teaching evaluation, one for research evaluation. He is also an elected member to the "Conseil National des Universités" (CNU) since October 2015.

Sophie Dabo-Niang offers expertise for Oreal's Award "Women in Science" since 2014.

10.1.7. Research Administration

Christophe Biernacki was "Délégué Scientifique Adjoint" of the Inria Lille center until June 2016. He is still member of the "Bureau du Comité des Projets" (BCP) of the Inria Lille center.

Sophie Dabo-Niang is the head of the MeQAME research team of Laboratory LEM-CNRS 9221.

Benjamin Guedj is a member of the scientific Council of the Laboratoire Paul Painlevé (maths department of the University of Lille).

Cristian Preda is a member of the Research Council of the University Lille 1.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Christophe Biernacki is the head of the M2 "Ingénierie Statistique et Numérique" http://mathematiques.univ-lille1.fr/Formation/ at University Lille 1.

Serge Iovleff is responsible for the Computer Science Licence semester.

Licence: Sophie Dabo-Niang, Probability, 24h, Lille 3, France

Licence: Serge Iovleff, Discrete Mathématics, 68H, Computer Science, IUT, Lille 1, France

Licence: Serge Iovleff, Linear Algebra 24H, Computer Science, IUT, Lille 1, France

Licence: Serge Iovleff, Analysis and Numerical Methods, 75H, Computer Science, IUT, Lille 1, France

Licence: Serge Iovleff, Graphs and languages, 36H, Computer Science, IUT, Lille 1, France

Licence: Serge Iovleff, Mathematical Modelisation, 38H, Computer Science, IUT, Lille 1, France

Licence: Serge Iovleff, Internships supervision, 15H, Computer Science, IUT, Lille 1, France

Licence: Serge Iovleff, Operational Research, 14H, Computer Science, IUT, Lille 1, France

Licence: Guillemette Marot, Biostatistics, 9h, L1, U. Lille Droit et Santé, France

Licence: Cristian Preda, Probability, 40h, L1, Polytech'Lille, France

Licence: Cristian Preda, Inferential Statistics, 50h, L1, Polytech'Lille, France

Licence: Vincent Vandewalle, Probability, 132h, L2, U. Lille 2, France

Licence: Vincent Vandewalle, Classification, 32h, L2, U. Lille 2, France

Licence: Vincent Vandewalle, Analysis, 24h, L2, U. Lille 2, France

Licence: Vincent Vandewalle, Machine Learning, 20h, L3, U. Lille 2, France

Master: Christophe Biernacki, coaching project, 10h, M1, U. Lille 1, France

Master: Christophe Biernacki, data analysis, 97.5h, M2, U. Lille 1, France

Master: Christophe Biernacki, coaching internship, 20h, M2, U. Lille 1, France

Master: Sophie Dabo-Niang, Advanced Statistics, 24h, Lille 3, France

Master: Sophie Dabo-Niang, Biostatistics Statistics, 40h, Master, Lille 1, France

Master: Sophie Dabo-Niang, Non-parametric Statistics, 24h, Master, UGB, Senegal

Master: Sophie Dabo-Niang, Spatial Statistics, 24h, Lille 3, France

Master: Benjamin Guedj, Statistical Learning: Theory and Algorithms, 18h, Lille 1, France

Master: Benjamin Guedj, Statistical Learning: Theory and Algorithms, 24h, Université Pierre & Marie Curie, France

Master: Benjamin Guedj, Statistical Learning: Theory and Algorithms, 30h, Institut de Statistique des Universités de Paris (ISUP), France

Master: Serge Iovleff, Object Oriented programming, 20H, Master Mathématiques Appliquées, Statistique - Ingénierie Mathématique, Lille 1, France

Master: Serge Iovleff, Probability and stochastic processes, AIMS (MBour-Sénégal)

Master: Guillemette Marot, Biostatistics, 44h, M1, U. Lille Droit et Santé, France

Master: Guillemette Marot, Coaching project, 12h, M1, U. Lille Droit et Santé, France

Master: Guillemette Marot, Supervised classification, 13h, M1, Polytech'Lille, France

Master: Cristian Preda, Data Analysis, 40h, M1, Polytech'Lille, France

Master: Cristian Preda, Biostatistics, 12h, M2, Polytech'Lille, France

Master: Vincent Vandewalle, Classification 60h, M1, U. Lille 1, France

Doctorat: Guillemette Marot, Data Analysis with R, 14h, U. Lille Droit et Santé, France

10.2.2. Supervision

PhD: Jérémie Kellner, "Gaussian processes and kernel methods", Université Lille 1, 12/2016, Alain Celisse.

PhD: Quentin Grimonprez, "Variable selection in high dimensional setting with correlation", Inria DGA & Université Lille 1, 12/2016, Guillemette Marot, Julien Jacques, Alain Celisse.

PHD: Florence Loingeville, "Modèle linéaire généralisé hiérarchique Gamma-Poisson pour le contrôle de qualité en microbiologie", Université Lille 1, 01/2016, Cristian Preda.

PhD: Mohamed Yayaha, Lille 3, Sophie Dabo-Niang and Aboubacar Amiri.

PhD: Aladji Bassene, Lille 3 & UGB (Sénégal), Sophie Dabo-Niang.

PhD in progress: Le Li, "PAC-Bayesian Online Clustering: theory and algorithms", iAdvize & Université d'Angers, 11/2014, Benjamin Guedj, Sébastien Loustau.

PhD in progress: Maxime Baelde, "Identification, localisation, séparation temps réel de sources sonores dans les flux audio multi-canaux", A-Volute, Inria & Université Lille 1, 01/2016, Christophe Biernacki.

PhD in progress: Anne-Lise Bedenel, "Appariement de descripteurs évoluant dans le temps", PIXEO, Inria & Université Lille 1, 06/2015, Christophe Biernacki, Laetitia Jourdan.

PhD in progress: Adrien Ehrhardt, "Modèles prédictifs pour données volumineuses et biaisées. Application à l'amélioration du scoring en risque crédit", CACF, Inria & Université Lille 1, 06/2016, Christophe Biernacki, Philippe Heinrich, Vincent Vandewalle.

PhD in progress: Maxime Brunin, "Early stopping rules in statistical learning", 09/2014, Christophe Biernacki, Alain Celisse.

PhD in progress: Emad Drwesh, Lille 3, Sophie Dabo-Niang, Jérôme Foncel.

PhD in progress: Mohamed Salem Ahmed, Lille 3, Sophie Dabo-Niang and Mohamed Attouch.

PhD in progress: H. Sarter, "Outils statistiques pour la sélection de variables et l'intégration de données cliniques et omiques : développement et application au registre EPIMAD", 12/2016, C. Gower, Guillemette Marot.

10.2.3. Juries

Christophe Biernacki participated as a reviewer to 5 PhD theses and 1 HdR committee, and as an examiner to 1 PhD thesis and 2 HdR committees. He also participated to 1 recruitment committee for a professor and was president of 1 recruitment committee for an assistant professor.

Alain Celisse has participated as an examiner to 1 PhD thesis.

Sophie Dabo-Niang has participated as an examiner to 4 PhD thesis.

Guillemette Marot was a member of two recruitment committees (MCU Univ. Nice, IE Univ. Lille). She was also an examiner to 1 PhD thesis.

Cristian Preda has participated as an examiner to 1 HdR committee.

Vincent Vandewalle has participated as an examiner to 1 PhD thesis.

10.3. Popularization

Christophe Biernacki has given about 10 talks during 2016 for institutions (Inria, universities, Ecole des Mines), companies and other related events. He gave also presentations towards students and industrials to the Xperium platform of the University of Lille about “Intelligence des données” (https://modal.lille.inria.fr/xperium/, http://learningcenters.nordpasdecalais.fr/innovation/fr/xperium). About 1,500 people came to this event during two years. He organized also a first short meeting in April 2016 in Lille for obtaining a feedback from company and academic users about the MASSICCC platform developped by the Modal and Select teams (https://massiccc.lille.inria.fr/#/). Here is the link towards this event: https://modal.lille.inria.fr/wikimodal/lib/exe/fetch.php?media=meeting_massiccc_7avril2016.pdf.

Sophie Dabo-Niang participates in the promotion of research among young children around a day of "Girls and Science, a light equation" organized in Lille (October 2016) and Senegal (Dakar, march 2016).

Benjamin Guedj has given a talk to high school students ("Terminale ISN") at Euratechnologies. The talk consisted in an overview of of machine learning impacts our everyday lives and how mathematicians contribute to learning in the big data era.

Vincent Vandewalle has given one presentation towards students to the Xperium platform of the University of Lille about “Intelligence des données” (https://modal.lille.inria.fr/xperium/, http://learningcenters.nordpasdecalais.fr/innovation/fr/xperium). He also has animated a formation on probabilities and statistics for middle School mathematics teachers through the Maison Pour la Science (http://www.maisons-pour-la-science.org/node/10641).

11. Bibliography

Major publications by the team in recent years

[1] S. ARLOT, A. CELISSE.*Segmentation of the mean of heteroscedastic data via cross-validation*, in "Statistics and Computing", 2010, vol. 21, p. 613–632.

[2] P. BATHIA, S. IOVLEFF, G. GOVAERT.*An R Package and C++ library for Latent block models: Theory, usage and applications*, in "Journal of Statistical Software", 2016, https://hal.archives-ouvertes.fr/hal-01285610.

[3] C. BIERNACKI, G. CELEUX, G. GOVAERT.*Exact and Monte Carlo Calculations of Integrated Likelihoods for the Latent Class Model*, in "Journal of Statistical Planning and Inference", 2010, vol. 140, p. 2991-3002, https://hal.archives-ouvertes.fr/hal-00554344.

[4] C. BIERNACKI, J. JACQUES.*A generative model for rank data based on an insertion sorting algorithm*, in "Computational Statistics and Data Analysis", 2013, vol. 58, p. 162-176 [*DOI :* 10.1016/J.CSDA.2012.08.008], https://hal.archives-ouvertes.fr/hal-00441209.

[5] C. BIERNACKI, J. JACQUES.*Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm*, in "Statistics and Computing", 2016, vol. 26, n^{o} 5, p. 929-943, https://hal.inria.fr/hal-01052447.

[6] A. CELISSE, J.-J. DAUDIN, L. PIERRE.*Consistency of maximum likelihood and variational estimators in stochastic block model*, in "Electronic Journal of Statistics", 2012, p. 1847–1899, http://projecteuclid.org/handle/euclid.ejs.

[7] M. GIACOFCI, S. LAMBERT-LACROIX, G. MAROT, F. PICARD.*Wavelet-based clustering for mixed-effects functional models in high dimension*, in "Biometrics", March 2013, vol. 69, n^o 1, p. 31-40 [*DOI :* 10.1111/J.1541-0420.2012.01828.X], http://hal.inria.fr/hal-00782458.

[8] J. JACQUES, C. PREDA.*Funclust: a curves clustering method using functional random variables density approximation*, in "Neurocomputing", 2013, vol. 112, p. 164-171, https://hal.archives-ouvertes.fr/hal-00628247.

[9] M. MARBAC, C. BIERNACKI, V. VANDEWALLE.*Model-based clustering of Gaussian copulas for mixed data*, in "Communications in Statistics - Theory and Methods", December 2016, https://hal.archives-ouvertes.fr/hal-00987760.

[10] V. VANDEWALLE, C. BIERNACKI, G. CELEUX, G. GOVAERT.*A predictive deviance criterion for selecting a generative model in semi-supervised classification*, in "Computational Statistics and Data Analysis", 2013, vol. 64, p. 220-236, https://hal.inria.fr/inria-00516991.

Publications of the year

Articles in International Peer-Reviewed Journal

[11] A. AMIRI, S. DABO-NIANG, M. YAHAYA.*Non-parametric recursive density estimation for spatial data*, in "Comptes Rendus Mathématique", 2016 [*DOI :* 10.1016/J.CRMA.2015.10.010], https://hal.inria.fr/hal-01425935.

[12] P. BATHIA, S. IOVLEFF, G. GOVAERT.*An R Package and C++ library for Latent block models: Theory, usage and applications*, in "Journal of Statistical Software", 2016, https://hal.archives-ouvertes.fr/hal-01285610.

[13] C. BIERNACKI, J. JACQUES.*Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm*, in "Statistics and Computing", 2016, vol. 26, n^o 5, p. 929-943, https://hal.inria.fr/hal-01052447.

[14] S. DABO-NIANG, S. GUILLAS, C. TERNYNCK.*More efficient kernel functional regression estimation with correlated errors*, in "Journal of Multivariate Analysis", 2016 [*DOI :* 10.1016/J.JMVA.2016.01.007], https://hal.inria.fr/hal-01425933.

[15] S. DABO-NIANG, A. MAINA, T. SOUBDHAN, H. OULD-BABA.*Predictive spatio-temporal model for spatially sparse global solar radiation data*, in "Energy", June 2016, vol. 111, p. 599–608 [*DOI :* 10.1016/J.ENERGY.2016.06.004], https://hal.inria.fr/hal-01425930.

[16] S. DABO-NIANG, C. TERNYNCK, A.-F. YAO.*Nonparametric prediction in the multivariate spatial context*, in "Journal of Nonparametric Statistics", 2016, vol. 28, n^o 2, p. 428-458 [*DOI :* 10.1080/10485252.2016.01.007], https://hal.inria.fr/hal-01425932.

[17] A. DERMOUNE, C. PREDA.*Parametrizations, fixed and random effects*, in "Journal of Multivariate Analysis", November 2016 [*DOI :* 10.1016/J.JMVA.2016.11.001], https://hal.archives-ouvertes.fr/hal-01424782.

[18] C. HERBAUX, E. BERTRAND, G. MAROT, C. ROUMIER, N. PORET, V. SOENEN, O. NIBOUREL, C. ROCHE-LESTIENNE, N. BROUCQSAULT, S. GALIÈGUE-ZOUITINA, E. BOYLE, G. FOUQUET, A. RENNEVILLE, S. TRICOT, F. MORSCHHAUSER, C. PREUDHOMME, B. QUESNEL, S. POULAIN, X.

LELEU.*BACH2 promotes indolent clinical presentation in Waldenström macroglobulinemia*, in "Oncotarget", 2016, https://hal.inria.fr/hal-01423307.

[19] J. JACQUES, C. RUCKEBUSCH.*Model-based co-clustering for hyperspectral images*, in "Journal of Spectral Imaging", 2016, https://hal.archives-ouvertes.fr/hal-01367941.

[20] M. MARBAC, C. BIERNACKI, V. VANDEWALLE.*Latent class model with conditional dependency per modes to cluster categorical data*, in "Advances in Data Analysis and Classification", June 2016, vol. 10, n° 2, p. 183–207 [*DOI :* 10.1007/S11634-016-0250-1], https://hal.archives-ouvertes.fr/hal-00950112.

[21] M. MARBAC, C. BIERNACKI, V. VANDEWALLE.*Model-based clustering of Gaussian copulas for mixed data*, in "Communications in Statistics - Theory and Methods", December 2016, https://hal.archives-ouvertes.fr/hal-00987760.

[22] P. MASSELOT, S. DABO-NIANG, F. CHEBANA, T. B. OUARDA.*Streamflow forecasting using functional regression*, in "Journal of Hydrology", April 2016, vol. 538, p. 754–766 [*DOI :* 10.1016/J.JHYDROL.2016.04.048], https://hal.inria.fr/hal-01425931.

[23] C. TERNYNCK, M. ALI BEN ALAYA, F. CHEBANA, S. DABO-NIANG, O. TAHA.*Streamflow hydrology classification using functional data analysis*, in "Journal of Hydrometeorology", 2016 [*DOI :* 10.1175/JHM-D-14-0200.1], https://hal.inria.fr/hal-01425934.

[24] L. YENGO, J. JACQUES, C. BIERNACKI, M. CANOUIL.*Variable Clustering in High-Dimensional Linear Regression: The R Package clere*, in "The R Journal", 2016, vol. 8, n° 1, p. 92-106, https://hal.archives-ouvertes.fr/hal-00940929.

Invited Conferences

[25] H. ALAWIEH, N. WICKER, C. BIERNACKI.*Projection under pairwise distance control*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2016, ERCIM 2016)", Séville, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01420681.

[26] C. BIERNACKI.*BigStat for Big Data: Big Data clustering through the BigStat SaaS platform*, in "Journée scientifique « Big Data & Data science »", Tunis, Tunisia, October 2016, https://hal.archives-ouvertes.fr/hal-01420650.

[27] C. BIERNACKI, M. BRUNIN, A. CELISSE.*Computation time/accuracy trade-off and linear regression*, in "9th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2016, ERCIM 2016)", Séville, Spain, December 2016, https://hal.archives-ouvertes.fr/hal-01420659.

[28] C. BIERNACKI, G. CASTELLAN, S. CHRETIEN, B. GUEDJ, V. VANDEWALLE.*Pitfalls in Mixtures from the Clustering Angle*, in "Working Group on Model-Based Clustering Summer Session", Paris, France, July 2016, https://hal.archives-ouvertes.fr/hal-01419755.

[29] C. BIERNACKI, A. LOURME.*Unifying Data Units and Models in Statistics: Focus on (Co-)Clustering*, in "Workshop on Model-based Clustering and Classification (MBC2)", Catania, Italy, September 2016, https://hal.archives-ouvertes.fr/hal-01420657.

[30] J. JACQUES, C. BIERNACKI.*Model-Based Co-clustering for Ordinal Data*, in "Working Group on Model-Based Clustering Summer Session", Paris, France, July 2016, https://hal.archives-ouvertes.fr/hal-01420648.

[31] J. JACQUES, C. BIERNACKI.*Model-based co-clustering for ordinal data*, in "23th Summer Working Group on Model-Based Clustering of the Department of Statistics of the University of Washington", Paris, France, July 2016, https://hal.inria.fr/hal-01383912.

Conferences without Proceedings

[32] M. BAELDE, C. BIERNACKI, R. GREFF.*A mixture model-based real-time audio sources classification method*, in "The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP2017", New Orleans, United States, March 2017, https://hal.archives-ouvertes.fr/hal-01420677.

[33] A.-L. BEDENEL, C. BIERNACKI, L. JOURDAN.*Appariement de descripteurs evoluant en temps Application a la comparaison d'assurance en ligne*, in "48èmes Journées de Statistique de la SFdS", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01420667.

[34] M. BRUNIN, C. BIERNACKI, A. CELISSE.*Compromis précision-temps de calcul appliqué au problèeme de détection de ruptures*, in "48èmes Journées de Statistique de la SFdS", Montpellier, France, May 2016, https://hal.archives-ouvertes.fr/hal-01420669.

[35] J. JACQUES, C. BIERNACKI.*Model-based co-clustering for ordinal data*, in "48èmes Journées de Statistique organisée par la Société Française de Statistique", Montpellier, France, 2016, https://hal.archives-ouvertes.fr/hal-01383927.

[36] J. JACQUES, C. RUCKEBUSCH.*Co-clustering for hyperspectral images*, in "6th International Conference in Spectral Imaging", Chamonix, France, July 2016, https://hal.archives-ouvertes.fr/hal-01383918.

Other Publications

[37] H. ALAWIEH, N. WICKER, C. BIERNACKI.*Projection under pairwise distance controls*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01420662.

[38] P. ALQUIER, B. GUEDJ.*An Oracle Inequality for Quasi-Bayesian Non-Negative Matrix Factorization*, August 2016, working paper or preprint, https://hal.inria.fr/hal-01251878.

[39] P. ALQUIER, B. GUEDJ.*Simpler PAC-Bayesian Bounds for Hostile Data*, October 2016, working paper or preprint, https://hal.inria.fr/hal-01385064.

[40] S. ARLOT, A. CELISSE, Z. HARCHAOUI.*A kernel multiple change-point algorithm via model selection*, March 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-00671174.

[41] A. CELISSE, B. GUEDJ.*Stability revisited: new generalisation bounds for the Leave-one-Out*, August 2016, working paper or preprint, https://hal.inria.fr/hal-01355365.

[42] A. CELISSE, G. MAROT, M. PIERRE-JEAN, G. RIGAILL.*New efficient algorithms for multiple change-point detection with kernels*, September 2016, working paper or preprint, https://hal.inria.fr/hal-01413230.

[43] S. DABO-NIANG, M. S. AHMED.*Generalized Functional Linear Models under Choice-Based Sampling*, July 2016, working paper or preprint, https://hal.inria.fr/hal-01345918.

[44] M. A. HASNAT, J. VELCIN, S. BONNEVAY, J. JACQUES.*Evolutionary clustering for categorical data using parametric links among multinomial mixture models*, March 2016, working paper or preprint, https://hal.inria.fr/hal-01204613.

[45] L. LI, B. GUEDJ, S. LOUSTAU.*Clustering en ligne : le point de vue PAC-bayésien*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01264934.

[46] L. LI, B. GUEDJ, S. LOUSTAU.*PAC-Bayesian Online Clustering*, January 2016, working paper or preprint, https://hal.inria.fr/hal-01264233.

[47] Y. B. SLIMEN, S. ALLIO, J. JACQUES.*Model-Based Co-clustering for Functional Data*, December 2016, working paper or preprint, https://hal.inria.fr/hal-01422756.

[48] V. VANDEWALLE, C. PREDA.*Clustering categorical functional data Application to medical discharge letters Medical discharge letters*, July 2016, Working Group on Model-Based Clustering Summer Session: Paris, July 17-23, 2016, Poster, https://hal.inria.fr/hal-01424950.

[49] V. VANDEWALLE.*Simultaneous dimension reduction and multi-objective clustering using probabilistic factorial discriminant analysis*, December 2016, CMStatistics 2016, https://hal.inria.fr/hal-01424965.

Project-Team NON-A

Non-Asymptotic estimation for online systems

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
CNRS

Ecole Centrale de Lille

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Optimization and control of dynamic systems

Table of contents

Project-Team NON-A

Creation of the Team: 2011 January 01, updated into Project-Team: 2012 July 01

Keywords:

Computer Science and Digital Science:
5.1.1. - Engineering of interactive systems
5.1.4. - Brain-computer interfaces, physiological computing
5.9.1. - Sampling, acquisition
5.9.2. - Estimation, modeling
5.10.3. - Planning
5.10.4. - Robot control
5.10.6. - Swarm robotics
6.1.1. - Continuous Modeling (PDE, ODE)
6.4.1. - Deterministic control
6.4.3. - Observability and Controlability
6.4.4. - Stability and Stabilization

Other Research Topics and Application Domains:
1.2. - Ecology
2.5.3. - Assistance for elderly
3.4.3. - Pollution
4.5. - Energy consumption
5.6. - Robotic systems
6.4. - Internet of things
6.6. - Embedded systems
7.1. - Traffic management
7.1.2. - Road traffic
7.2.1. - Smart vehicles

1. Members

Research Scientists
Denis Efimov [Inria, Researcher, HDR]
Thierry Floquet [CNRS, Senior Researcher]
Andrey Polyakov [Inria, Researcher]
Alban Quadrat [Inria, Senior Researcher, HDR]
Rosane Ushirobira [Inria, Researcher, HDR]
Gang Zheng [Inria, Researcher, HDR]

Faculty Members
Jean-Pierre Richard [Team leader, Ecole Centrale de Lille, Professor, HDR]
Lotfi Belkoura [Univ. Lille 1, Professor, HDR]
Wilfrid Perruquetti [Ecole Centrale de Lille, Professor, HDR]

Technical Staff
Roudy Dagher [Inria]

PhD Students

Hafiz Ahmed [Inria, until Sep 2016]
Maxime Feingesicht [Ecole Centrale de Lille]
Zohra Kader [Univ. Lille 1]
Tatiana Kharkovskaia [ITMO University]
Kokou Langueh [Ecole Centrale de Lille]
Francisco Lopez Ramirez [Inria]
Nadhynee Martinez Fonseca [National Polytechnic Institute of Mexico, from Sep 2016]
Gabriele Perozzi [ONERA]
Guillaume Rance [Safran Electronics & Defense, granted by CIFRE]
Zilong Shao [EC Lille, until May 2016]
Haik Jan Silm [Ecole Centrale de Lille, from Oct 2016]
Jijju Thomas [Ecole Centrale de Lille, from Nov 2016]
Yue Wang [Ecole Centrale de Lille, from Feb 2016]

Post-Doctoral Fellows

Stanislav Aranovskiy [Inria]
Mohamed Yacine Bouzidi [Ecole Centrale de Lille, Faculty Member]
Lucien Etienne [Inria, from Feb 2016]

Visiting Scientists

Arie Levant [Inria, Researcher, until Jan 2016]
Emilia Fridman [Ecole Centrale de Lille, until Jul 2016]
Leonid Fridman [UNAM Mexico, until Jul 2016]
Jaime Moreno Perez [Universidad Nacional Autonoma de Mexico, from Jun 2016 until Jul 2016]
Juan Rueda Escobedo [UNAM Mexico, from Oct 2016 until Nov 2016]
Tonametl Sanchez Ramirez [UNAM Mexico, until Nov 2016]
Konstantin Zimenko [ITMO University St Petersburg, until Oct 2016]

Administrative Assistant

Corinne Jamroz [Inria]

Others

Boussad Abci [Inria, from Feb 2016 until May 2016]
Jean-Pierre Barbot [ENSEA, Professor, HDR]
Maelle Das Neves [Inria, Oct 2016]
Baihui Du [Ecole Centrale de Lille, from May 2016 until Jul 2016]
Olivier Gibaru [ENSAM, Faculty Member, HDR]
Laurentiu Hetel [CNRS, Researcher, HDR]
Cédric Join [Univ. Nancy I, Faculty Member, HDR]
Paul Lesur [Inria, from May 2016 until Jul 2016]
Mamadou Mboup [Univ. Reims, Faculty Member, HDR]
Stephane Thiery [ENSAM]
Djahid Rabehi [Ecole Centrale de Lille, from Apr 2016 until Jul 2016]
Samer Riachy [ENSEA, Faculty Member]
Arthur Varre [Inria, Dec 2016]

2. Overall Objectives

2.1. Objectives

For engineers, a wide variety of information cannot be directly obtained through measurements. Some parameters (constants of an electrical actuator, delay in a transmission, etc.) or internal variables (robot's posture, torques applied to a robot, localization of a mobile robot, etc.) are unknown or unmeasured. In addition, usually the signals from sensors are distorted and tainted by measurement noises. In order to simulate, to control or to supervise processes, and to extract information conveyed by the signals, one has to estimate parameters or variables.

Estimation techniques are, under various guises, present in many parts of control, signal processing and applied mathematics. Such an important area gave rise to a huge international literature. From a general point of view, the performance of an estimation algorithm can be characterized by three indicators:

- The computation time (the time needed to obtain the estimation). Obviously, the estimation algorithms should have as small as possible computation time in order to provide fast, real-time, on-line estimations for processes with fast dynamics (for example, a challenging problem is to make an Atomic Force Microscope work at GHz rates).
- The algorithm complexity (the easiness of design and implementation). Estimation algorithms should have as low as possible algorithm complexity, in order to allow an embedded real-time estimation (for example, in networked robotics, the embedded computation power is limited and can be even more limited for small sensors/actuators devices). Another question about complexity is: can an engineer appropriate and apply the algorithms? For instance, an algorithm application is easier if the parameters have a physical meaning w.r.t. the process under study.
- The robustness. The estimation algorithms should exhibit as much as possible robustness with respect to a large class of measurement noises, parameter uncertainties, discretization steps and other issues of numerical implementation. A complementary point of view on robustness is to manage a compromise between existence of theoretical proofs versus universalism of the algorithm. In the first case, the performance is guaranteed in a particular case (a particular control designed for a particular model). In the second case, an algorithm can be directly applied in “most of the cases", but it may fail in few situations.

Within the very wide area of estimation, *Non-A* addresses 3 particular theoretical challenges (see the upper block “Theory” of Figure 1):

1) Design annihilators for some general class of perturbations;

2) Estimate on-line the derivatives of a signal;

3) Control without sophisticated models.

All of them are connected with the central idea of designing or exploiting algorithms with the finite-time convergence property. In particular, the *non-asymptotic* estimation techniques (numerical differentiation, finite-time differentiators or observers) constitute a central objective of the project, explaining the name *Non-Asymptotic estimation for on-line systems*. Below, these 3 challenges will be shortly described in relation to the above indicators.

The researches developed by *Non-A* are within the continuity of the project-team *ALIEN* in what concerns the *algebraic tools* that are developed for finite-time estimation purposes. However, *Non-A* also aims at developing complementary estimation techniques, still aiming at the finite-time performance but based on the so-called *higher-order sliding mode* algorithms, interval estimation techniques and, as well as, fixed-time algorithms.

Non-A also wants to confront these theoretical challenges with some application fields (shown on the bottom of Figure 1): Networked robots, Nano/macro machining, Multicell chopper, *i*-PID for industry. Today, most of our effort (*i.e.*, engineering staff) is devoted to the first item, according to the theme ‘Internet of Things’ promoted by Inria in its Strategic Plan for the Lille North-Europe research center. Indeed, WSNR (Wireless Sensor and Robot Networks) integrate mobile nodes (robots) that extends various aspects of the sensor network.

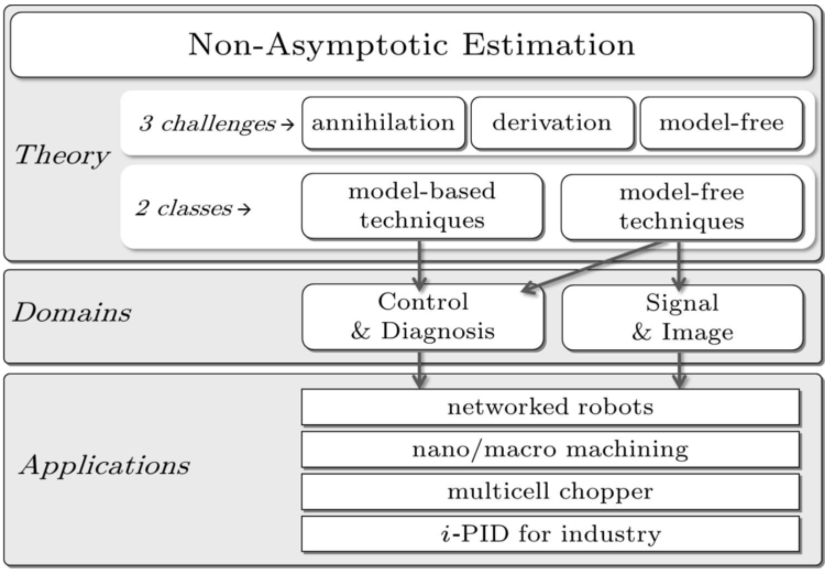


Figure 1. Non-A is a method-driven project, centered around non-asymptotic estimation techniques (i.e. providing estimates in finite-time), and connected to applications.

2.2. Members complementarity

The members of the Non-A project work in different places: Lille, Cergy, Reims and Nancy. They share a common algebraic tool and the non-asymptotic estimation goal, which constitute the natural kernel of the project. Each of them contributes to both theoretical and applied sides of the global project. The following table draws up a scheme of some of their specialities.

	Upstream Researches	*Application Fields*
Reims CReSTIC	Signal - Numerical analysis	De-noising - Demodulation - Biomedical signal processing
Cergy Quartz	Nonlinear observers - Hybrid systems	Cryptography - Multi-cell chopper/converter
Lille ENSAM	Applied mathematics	High performance machining - Precision sensors, AFM [0]
Lille CRIStAL	Delay systems - Nonlinear control - Observers (finite-time/unknown input)	Magnetic bearings - Friction estimation - Networked control - Robotics
Nancy CRAN	Diagnosis - Control - Signal	Industrial processes - Signal & image processing

3. Research Program

3.1. General annihilators

[0] Atomic Force Microscope, for which fast filtering is required

Estimation is quite easy in the absence of perturbations. It becomes challenging in more realistic situations, faced to measurement noises or other unknown inputs. In our works, as well as in the founding text of *Non-A*, we have shown how our estimation techniques can successfully get rid of perturbations of the so-called *structured* type, which means the ones that can be annihilated by some linear differential operator (called the annihilator). *ALIEN* already defined such operators by integral operators, but using more general convolution operators is an alternative to be analyzed, as well as defining the "best way to kill" perturbations. Open questions are:

OQ1) Does a normal form exist for such annihilators?

OQ2) Or, at least, does there exist an adequate basis representation of the annihilator in some adequate algebra?

OQ3) And lastly, can the annihilator parameters be derived from efficient tuning rules?

The two first questions will directly impact Indicators 1 (time) and 2 (complexity), whereas the last one will impact indicator 3 (robustness).

3.2. Numerical differentiation

Estimating the derivative of a (noisy) signal with a sufficient accuracy can be seen as a key problem in domains of control and diagnosis, as well as signal and image processing. At the present stage of our research, the estimation of the n-th order time derivatives of noisy signals (including noise filtering for $n = 0$) appears as a common area for the whole project, either as a research field, or as a tool that is used both for model-based and model-free techniques. *One of the open questions is about the robustness issues (Indicator 3) with respect to the annihilator, the parameters and the numerical implementation choices.*

Two classes of techniques are considered here (**Model-based** and **Model-free**), both of them aiming at non-asymptotic estimation.

In what we call *model-based techniques*, the derivative estimation is regarded as an observation problem, which means the software-based reconstruction of unmeasured variables and, more generally, a left inversion problem [0]. This involves linear/homogeneous/nonlinear state models, including ordinary equations, systems with delays, hybrid systems with impulses or switches [0], which still has to be exploited in the finite-time and fixed-time context. Power electronics is already one of the possible applications.

Model-free techniques concern the works initiated by *ALIEN*, which rely on the only information contained in the output signal and its derivatives. The corresponding algorithms rely on our algebraic annihilation viewpoint. *One open question is: How to provide an objective comparison analysis between Model-based and Model-free estimation techniques? For this, we will only concentrate on Non-Asymptotic ones. This comparison will have to be based on the three Indicators 1 (time), 2 (complexity) and 3 (robustness).*

3.3. Model-free control

Industry is keen on simple and powerful controllers: the tuning simplicity of the classical PID controller explains its omnipresence in industrial control systems, although its performances drop when working conditions change. The last challenge we consider is to define control techniques which, instead of using sophisticated models (the development of which may be expensive), use the information contained in the output signal and its estimated derivatives, which can be regarded as "signal-based" controllers. *Such design should take into account the Indicators 1 (time), 2 (complexity) and 3 (robustness).*

[0] Left invertibility deals with the question of recovering the full state of a system ("observation") together with some of its inputs ("unknown input observers"), and also refers to algebraic structural conditions.

[0] Note that hybrid dynamical systems (HDS) constitute an important field of investigation since, in this case, the discrete state can be considered as an unknown input.

3.4. Applications

Keeping in mind that we will remain focused at developing and applying fundamental methods for non-asymptotic estimation, we intend to deal with 4 main domains of application (see the lower part of Figure 1). The Lille context offers interesting opportunities in WSAN (wireless sensor and actuator networks and, more particularly, networked robots) at Inria, as well as nano/macro machining at ENSAM. A power electronics platform will be developed in ENSEA Cergy. Last, in contact with companies, several grants, patents and collaborations are expected from the applications of $i-$PID. Each of these four application domains was presented in the *Non-A* proposal:

- Networked robots, WSAN [Lille]
- Nano/macro machining [Lille]
- Multicell chopper [Lille and Cergy]
- i-PID for industry

In the present period, we choose to give a particular focus to the first item (Networked robots), which already received some development. It can be considered as the objective 4.

4. Application Domains

4.1. Robots and networked systems

Inria Lille and team FUN are hosting an "equipment of excellence", named FIT-IoT lab. It gives a remote access to thousands of wireless sensors to be connected with hundreds of mobile robots. Today, many sensor scenarios are available, with few robot testbeds.

The package SLIM, developed by Non-A under ROS (Robot Operating System) with the support of an Inria ADT, aims at contributing to this environment. The self deployment of autonomous groups of mobile robots in an unknown and variable environment is a next step for IoT-lab, involving localization, path planning and robust control problems. Our ROS package SLIM aims at combining various algorithms developed by Non-A (localization, path planning, robust control). It should also offer a software library for multi-robot including: optimal local planner based on flatness; plugin for communication between different ROS cores; module Multi-Mapping for robot cooperation; plugin for YEI IMU.

4.2. Living systems: ecological monitoring, modelling, estimation and identification of biological systems, human-computer interaction

Modelling, estimation or detection for living is difficult because such systems cannot be isolated from external influences. Using our numerical differentiation tools, together with modelling techniques, we want to study the following four applications:

- *Biosensing*: Unlike classical approaches deploying physical sensors, biological systems can be used as living sensors. The marine biology lab EPOC (CNRS, Bordeaux) has developed underwater sensors for bivalve molluscs (such as oysters) measuring and sending through RGPS the opening gap between the two valves. We want to use it for water quality monitoring by either identifying oyster's rythm I/O models or by using our differentiation tools. Spawning detection is also considered (ANR WaQMoS).
- *Human-Computer Interaction*: Reduction of the latency between the human input and the system visual response in HCI (ANR TurboTouch). To do that, a simple forecasting algorithm for latency compensation in indirect interaction using a mouse has to be developed based on differentiators.
- *Smart bracelet*: Design a dynamical model for the GSR and for the development of an online algorithm making the GSR signal independent of the user movements. Most resulting computations should be embedded into the bracelet. Collaboration with NEOTROPE (start-up developing a bracelet intended for strong human emotion detection).
- *Microbial populations*: Real-time control of synthetic microbial communities (Inria Project Lab, COSY, under evaluation).

4.3. Turbulent flow control for aircrafts and vehicles

Non-A is active in a Regional consortium gathering micro-technologies (ONERA, IEMN, LAMIH, LML and PPrime lab, Univ. of Poitiers) which aims at developing methods for active control of separated flows (ContrATech subprogram of CPER ELSAT).

Aerodynamic losses are believed to be a major source of energy wastage for a vehicle at speeds higher than 50 km/h. Optimization of the vehicles shapes has reached its limit and such a passive control approach cannot deal with unsteady incoming flow. Similarly, in aeronautics, controlling boundary layer airflow could reduce stall drastically. In such contexts, active control strategies (air blowers, hot film sensors, etc.) are very attractive. But the natural phenomena ruling turbulent flows lead to highly nonlinear and infinite-dimension dynamics. Till now, researchers use either nonlinear PDEs (Navier-Stokes equations) allowing for analysis but improper for control design or unrealistic linear finite-dimension models for classical − but non robust − control. Non-A first wants to propose a model with intermediate complexity (bilinear with time delays, "grey-box" identification on experimental data) and then develop model-based sliding mode and optimal control algorithms.

4.4. Industry and society: i-PID for industry and society, mechatronics (Safran)

- Industry is keen on simple and powerful controllers. The tuning simplicity of the classical PID controller explains its omnipresence in industrial control systems, although its performances drop when the working conditions change. AL.I.E.N SAS was created in 2011 as a spin-off of the Inria project ALIEN, which gave rise to Non-A, working on algebraic estimation and i-PID controller (i.e., using algebraic estimation of the perturbations and apply a simple PID control on some "ultra-local" model). These control technique uses the information contained in the output signal and its estimated derivatives, which can be regarded as "signal-based" controllers. Model-free control technique has been applied in many different domains (electronics, hydroelectric power, etc.).

  Recent research is focused on traffic control and biology. The quality of traffic control laws depends on a good knowledge of the highway characteristics, especially the critical density and the free-flow speed, which are unfortunately most difficult to estimate in real time. Therefore, we aim at developing an algorithm which shows the possibility to control the traffic without the knowledge of density and free-flow speed.
- A collaboration with the Safran Electronics & Defense company has been developed (CIFRE PhD thesis) on the parametric stabilization of gyrostabilized platforms. To do that, we first aim at developing new symbolic-numeric methods for the standard H_∞-loop shaping design problem for models of gyrostabilized platforms in terms of the physical parameters (masses, inertia, etc.) considered as unknown/slowly varying parameters. Using Non-A techniques for the estimation of the physical parameters, we then want to develop new embeddable and adaptive controllers for the robust stabilization of gyrostabilized platforms.

5. Highlights of the Year

5.1. Highlights of the Year

UCoCoS

The H2020 project UCoCoS (Understanding and Controlling of Complex Systems, supervisors: W. Michiels, J.-P. Richard, H. Nijmeijer, 2016-2020) has started effectively this year: kick-off meeting in Eindhoven in March and, at the end of this year, recruitment of the 6 PhD students (including 4 jointly with Lille: H. Silm, J. Thomas, D. Dileep, Q. Voortman) in the 3 hosting institutions.

5.1.1. Awards

D. Efimov is Outstanding IEEE TAC reviewer.

6. New Software and Platforms

6.1. Blimp

FUNCTIONAL DESCRIPTION

Scientific research and development on the control of autonomous airship have shown a significant growth in recent years. New applications appear in the areas such as freight carrier, advertising, monitoring, surveillance, transportation, military and scientific research. The control of autonomous airship is a very important problem for the aerial robots research.

The development of Blimp by Non-A is used for experimentation and demonstration of controlling algorithms. The blimp is required to provide some environment information and status of itself, such as surveillance video of surrounding environment, gesture of blimp, altitude of blimp. With these basic information, one could localize blimp with certain algorithm (visual SLAM for example) or implement one controller in order to improve the stability and maneuverability of blimp.

- Contact: Jean-Pierre Richard
- URL: https://bil.inria.fr/fr/software/view/2279/tab

6.2. ControlHub

The driving idea is to interconnect a group of actors (researchers, engineers, etc.) around a control problem and grant them remote access to existing experimental facilities, thus allowing them to verify their theoretical results online, and finally share them with the project members.

The platform architecture relies on three key principles:

- Problem centric: The control problem to be solved is the core project, whereas the software resources, tools and online experiments are web services available to support experimental verification of the solutions.
- Separation of concerns: setup and maintenance of experiment facilities, installation of software tools, problem formulation and theoretical analysis, etc.
- Resource sharing: software packages, experimental facilities, open problems.
- Contact: R. Dagher, A. Polyakov, J.-P. Richard
- URL: https://bil.inria.fr/fr/software/view/2830/tab

6.3. SLIM

FUNCTIONAL DESCRIPTION

Multi-robots cooperation can be found as an application in many domains of science and technology: manufacturing, medical robotics, personal assistance, military/security and spatial robots. The market of robots is quickly developing and its capacity is continuously growing. Concerning cooperation of mobile multi-robots, 3 key issues have to be studied: Localization, path planning and robust control, for which Non-A team has worked and proposed new algorithms. Due to the ADT SLIM, we implement our algorithms (localization, path planning and robust control) and integrate them into ROS (Robotic Operating System) as a package, named SLIM.

- Contact: G. Zheng
- URL: https://bil.inria.fr/fr/software/view/2278/tab

7. New Results

7.1. Homogeneity Theory

Homogeneity is one of the tools we develop for finite-time convergence analysis. In 2016 this concept has received various improvements:

- Frequency domain approach to analysis of homogeneous nonlinear systems [85], [46]:

  Analysis of feedback sensitivity functions for implicit Lyapunov function-based control system is developed. The Gang of Four and loop transfer function are considered for practical implementation of the control via frequency domain control design. The effectiveness of this control scheme is demonstrated on an illustrative example of roll control for a vectored thrust aircraft.

- Homogeneous distributed parameter systems [72], [32]:

  A geometric homogeneity is introduced for evolution equations in a Banach space. Scalability property of solutions of homogeneous evolution equations is proven. Some qualitative characteristics of stability of trivial solution are also provided. In particular, finite-time stability of homogeneous evolution equations is studied. Classical theorems on existence and uniqueness of solutions of nonlinear evolution equations are revised. A universal homogeneous feedback control for a finite-time stabilization of linear evolution equation in a Hilbert space is designed using homogeneity concept. The design scheme is demonstrated for distributed finite-time control of heat and wave equations.

- Robustness of Homogenous Systems:

  – [93], [36] The problem of stability robustness with respect to time-varying perturbations of a given frequency spectrum is studied applying homogeneity framework. The notion of finite-time stability over time intervals of finite length, i.e. short-finite-time stability, is introduced and used for that purpose. The results are applied to demonstrate some robustness properties of the three-tank system.

  – The uniform stability notion for a class of nonlinear time-varying systems is studied in [35] using the homogeneity framework. It is assumed that the system is weighted homogeneous considering the time variable as a constant parameter, then several conditions of uniform stability for such a class of systems are formulated. The results are applied to the problem of adaptive estimation for a linear system.

  – Robustness with respect to delays is discussed in [84], [45] for homogeneous systems with negative degree. It is shown that if homogeneous system with negative degree is globally asymptotically stable at the origin in the delay-free case then the system is globally asymptotically stable with respect to a compact set containing the origin independently of delay. The possibility of applying the result for local analysis of stability for not necessary homogeneous systems is analyzed. The theoretical results are supported by numerical examples.

- Finite-time and Fixed-time Control and Estimation:

  – [61], [46] A switched supervisory algorithm is proposed, which ensures fixed-time convergence by commutation of finite-time or exponentially stable homogeneous systems of a special class, and a finite-time convergence to the origin by orchestrating among asymptotically stable systems. A particular attention is paid to the case of exponentially stable systems. Finite-time and fixed-time observation problem of linear multiple input multiple output (MIMO) control systems is studied. The nonlinear dynamic observers , which guarantee convergence of the observer states to the original system state in a finite and a fixed (defined a priori) time, are studied. Algorithms for the observers parameters tuning are also developed.

- [16] This paper focuses on the design of fixed-time consensus for multiple unicycle-type mobile agents. A distributed switched strategy, based on local information, is proposed to solve the leader-follower consensus problem for multiple nonholonomic agents in chained form. The switching times and the prescribed convergence time are explicitly given regardless of the initial conditions. Simulation results highlight the efficiency of the proposed method.

- Discretization of Homogeneous Systems:
  - [63] Sufficient conditions for the existence and convergence to zero of numeric approximations to solutions of asymptotically stable homogeneous systems are obtained for the explicit and implicit Euler integration schemes. It is shown that the explicit Euler method has certain drawbacks for the global approximation of homogeneous systems with non-zero degrees, whereas the implicit Euler scheme ensures convergence of the approximating solutions to zero.
  - [69] The known results on asymptotic stability of homogeneous differential inclusions with negative homogeneity degrees and their accuracy in the presence of noises and delays are extended to arbitrary homogeneity degrees. Discretization issues are considered, which include explicit and implicit Euler integration schemes. Computer simulation illustrates the theoretical results.
- Multi-Homogeneity and differential inclusions:
  - The notion of homogeneity in the bi-limit from is extended in [21] to local homogeneity and then to homogeneity in the multi-limit. The converse Lyapunov/Chetaev theorems on (homogeneous) system instability are obtained. The problem of oscillation detection for nonlinear systems is addressed. The sufficient conditions of oscillation existence for systems homogeneous in the multi-limit are formulated. The proposed approach estimates the number of oscillating modes and the regions of their location. Efficiency of the technique is demonstrated on several examples.
  - In [94], the notion of geometric homogeneity is extended for differential inclusions. This kind of homogeneity provides the most advanced coordinate-free framework for analysis and synthesis of nonlinear discontinuous systems. The main qualitative properties of continuous homogeneous systems are extended to the discontinuous setting: the equivalence of the global asymptotic stability and the existence of a homogeneous Lyapunov function; the link between finite-time stability and negative degree of homogeneity; the equivalence between attractivity and asymptotic stability are among the proved results.

7.2. Algebraic Technique For Estimation

- Time parameter estimation for a sum of sinusoidal waveform signals [39]:

  A novel algebraic method is proposed to estimate amplitudes, frequencies, and phases of a biased and noisy sum of complex exponential sinusoidal signals. The resulting parameter estimates are given by original closed formulas, constructed as integrals acting as time-varying filters of the noisy measured signal. The proposed algebraic method provides faster and more robust results, compared with usual procedures. Some computer simulations illustrate the efficiency of our method.
- Algebraic estimation via orthogonal polynomials [80]:

  Many important problems in signal processing and control engineering concern the reconstitution of a noisy biased signal. For this issue, we consider the signal written as an orthogonal polynomial series expansion and we provide an algebraic estimation of its coefficients. We specialize in Hermite polynomials. On the other hand, the dynamical system described by the noisy biased signal may be given by an ordinary differential equation associated with classical orthogonal polynomials. The signal may be recovered through the coefficients identification. As an example, we illustrate our algebraic method on the parameter estimation in the case of Hermite polynomials.

- An effective study of the algebraic parameter estimation problem [105]:

  Within the algebraic analysis approach, we first give a general formulation of the algebraic parameter estimation for signals which are defined by ordinary differential equations with polynomial coefficients such as the standard orthogonal polynomials (Chebyshev, Jacobi, Legendre, Laguerre, Hermite, ... polynomials). We then show that the algebraic parameter estimation problem for a truncated expansion of a function into an orthogonal basis of L^2 defined by orthogonal polynomials can be studied similarly. Then, using symbolic computation methods such as Gröbner basis techniques for (noncommutative) polynomial rings, we first show how to compute ordinary differential operators which annihilate a given polynomial and which contain only certain parameters in their coefficients. Then, we explain how to compute the intersection of the annihilator ideals of two polynomials and characterize the ordinary differential operators which annihilate a first polynomial but not a second one. These results are implemented in the NON-A package built upon the OREMODULES software.

7.3. Set-Theoretic Methods of Control, Observer Design and Estimation

- Interval Observers:
  - [19] New design of interval observers for continuous-time systems with discrete-time measurements is proposed. For this purpose new conditions of positivity for linear systems with sampled feedbacks are obtained. A sampled-data stabilizing control is synthesized based on provided interval estimates. Efficiency of the obtained solution is demonstrated on examples.
  - [66] The problem of interval state estimation is studied for systems described by parabolic Partial Differential Equations (PDEs). The proposed solution is based on a finite-element approximation of PDE, with posterior design of an interval observer for the obtained ordinary differential equation. The interval inclusion of the state function of PDE is obtained using the estimates on the error of discretization. The results are illustrated by numerical experiments with an academic example.
  - [18] New interval observers are designed for linear systems with time-varying delays in the case of delayed measurements. Interval observers employ positivity and stability analysis of the estimation error system, which in the case of delayed measurements should be delay-dependent. New delay-dependent conditions of positivity for linear systems with time-varying delays are introduced. Efficiency of the obtained solution is demonstrated on examples.
  - [22] Interval state observers provide an estimate on the set of admissible values of the state vector at each instant of time. Ideally, the size of the evaluated set is proportional to the model uncertainty, thus interval observers generate the state estimates with estimation error bounds, similarly to Kalman filters, but in the deterministic framework. Main tools and techniques for design of interval observers are reviewed in this tutorial for continuous-time, discrete-time and time-delayed systems.
  - [43] investigates the problem of observer design for a general class of linear singular time-delay systems, in which the time delays are involved in the state, the output and the known input (if there exists). The involvement of the delay could be multiple which however is rarely studied in the literature. Sufficient conditions are proposed which guarantees the existence of a Luenberger-like observer for the general system.
  - In [90] an interval observer is proposed for on-line estimation of differentiation errors in some class of high-order differentiators (like a high-gain differentiator, or homogeneous nonlinear differentiator, or super-twisting differentiator). The results are verified and validated on the telescopic link of a robotic arm for forestry applications in which the mentioned approaches are used to estimate the extension velocity while the interval observer gives bounds to this estimation.

  - The problem of interval observer design is studied in [87] for a class of linear hybrid systems. Several observers are designed oriented on different conditions of positivity and stability for estimation error dynamics. Efficiency of the proposed approach is demonstrated by computer experiments for academic and bouncing ball systems.
  - The problem of estimation of sequestered parasites Plasmodium falciparum in malaria, based on measurements of circulating parasites, is addressed in [60]. It is assumed that all (death, transition, recruitment and infection) rates in the model of a patient are uncertain (just intervals of admissible values are given) and the measurements are subject to a bounded noise, then an interval observer is designed. Stability of the observer can be verified by a solution of LMI. The efficiency of the observer is demonstrated in simulation.
- Observer design:
  - [81] presents a new approach for observer design for a class of nonlinear singular systems which can be transformed into a special normal form. The interest of the proposed form is to facilitate the observer synthesis for the studied nonlinear singular systems. Necessary and sufficient geometrical conditions are deduced in order to guarantee the existence of a diffeomorphism which transforms the studied nonlinear singular systems into the proposed normal form.
  - In [38], we investigate the estimation problem for a class of partially observable nonlinear systems. For the proposed Partial Observer Normal Form (PONF), necessary and sufficient conditions are deduced to guarantee the existence of a change of coordinates which can transform the studied system into the proposed PONF. Examples are provided to illustrate the effectiveness of the proposed results.
  - [71] deals with the problem of finite-time and fixed-time observation of linear multiple input multiple output (MIMO) control systems. The nonlinear dynamic observers , which guarantee convergence of the observer states to the original system state in a finite and a fixed (defined a priori) time, are studied. Algorithms for the observers parameters tuning are also developed. The theoretical results are illustrated by numerical examples.
  - [44] Sliding mode control design for linear systems with incomplete and noisy measurements of the output and additive/multiplicative exogenous disturbances is studied. A linear minimax observer estimating the system's state with minimal worst-case error is designed. An algorithm, generating continuous and discontinuous feedbacks, which steers the state as close as possible to a given sliding hyperplane in finite time, is presented. The optimality (sub-optimality) of the designed feedbacks is proven for the case of bounded noises and additive (multiplicative) disturbances of L_2-class.
  - [37] deals with the design of a robust control for linear systems with external disturbances using a homogeneous differentiator-based observer based on a implicit Lyapunov function approach. Sufficient conditions for stability of the closed-loop system in the presence of external disturbances are obtained and represented by linear matrix inequalities. The parameter tuning for both controller and observer is formulated as a semi-definite programming problem with linear matrix inequalities constraints. Simulation results illustrate the feasibility of the proposed approach and some improvements with respect to the classic linear observer approach.
  - The problem studied in [17] is one of improving the performance of a class of adaptive observer in the presence of exogenous disturbances. The H^∞ gains of both a conventional and the newly proposed sliding-mode adaptive observer are evaluated, to assess the effect of disturbances on the estimation errors. It is shown that if the disturbance is “matched” in the plant equations, then including an additional sliding-mode feedback injection term, dependent on the plant output, improves the accuracy of observation.

  - In [95], we consider the classical reaching problem of sliding mode control design, that is to find a control law which steers the state of a Linear Time-Invariant (LTI) system towards a given hyperplane in a finite time. Since the LTI system is subject to unknown but bounded disturbances we apply the minimax observer which provides the best possible estimate of the system's state. The reaching problem is then solved in observer's state space by constructing a feedback control law. The cases of discontinuous and continuous admissible feedbacks are studied. The theoretical results are illustrated by numerical simulations.
- Estimation and Identification:
  - The problem of output control for linear uncertain system with external perturbations is studied in [77]. It is assumed that the output available for measurements is the higher order derivative of the state only (acceleration for a second order plant), which is also corrupted by noise. Then via series of integration an identification algorithm is proposed for identification of values of all parameters and unknown initial conditions for the state vector. Finally, two control algorithms are developed, adaptive and robust, providing boundedness of trajectories for the system. Efficiency of the obtained solutions is demonstrated by numerical experiments.
  - [24] focuses on the problem of velocity and position estimation. A solution is presented for a class of oscillating systems in which position, velocity and acceleration are zero mean signals. The proposed scheme considers that the dynamic model of the system is unknown. Only noisy acceleration measurements, that may be contaminated by zero mean noise and constant bias, are considered to be available. The proposal uses the periodic nature of the signals obtaining finite-time estimations while tackling integration drift accumulation.
  - In [41], we investigate the problem of simultaneous state and parameter estimation for a class of nonlinear systems which can be transformed into an output depending normal form. A new and simple adaptive observer for such class of systems is presented. Sufficient condition for the existence of the proposed observer is derived. A concrete application is given in order to highlight the effectiveness of the proposed result.
  - In [75], the problem of time-varying parameter identification is studied. To this aim, an identification algorithm is developed in order to identify time-varying parameters in a finite-time. The convergence proofs are based on a notion of finite-time stability over finite intervals of time, i.e. Short-finite-time stability; homogeneity for time-varying systems; and Lyapunov function approach. The algorithm asks for a condition over the regressor term which is related to the classic identifiability condition corresponding to the injectivity of such a term. Simulation results illustrate the feasibility of the proposed algorithm.

7.4. Stability, Stabilization, Synchronization

- Input-to-state stability:
  - Supported by a novel field definition and recent control theory results, a new method to avoid local minima is proposed in [25]. It is formally shown that the system has an attracting equilibrium at the target point, repelling equilibriums in the obstacles centers and saddle points on the borders. Those unstable equilibriums are avoided capitalizing on the established Input-to-State Stability (ISS) property of this multistable system. The proposed modification of the PF method is shown to be effective by simulation for a two variables integrator and then applied to an unicycle-like wheeled mobile robots which is subject to additive input disturbances.
  - [62] Motivated by the problem of phase-locking in droop-controlled inverter-based microgrids with delays, the recently developed theory of input-to-state stability (ISS) for multistable systems is extended to the case of multistable systems with delayed dynamics. Sufficient conditions for ISS of delayed systems are presented using Lyapunov-Razumikhin

functions. It is shown that ISS multistable systems are robust with respect to delays in a feedback. The derived theory is applied to two examples. First, the ISS property is established for the model of a nonlinear pendulum and delay-dependent ro-bustness conditions are derived. Second, it is shown that, under certain assumptions, the problem of phase-locking analysis in droop-controlled inverter-based microgrids with delays can be reduced to the stability investigation of the nonlinear pendulum. For this case, corresponding delay-dependent conditions for asymptotic phase-locking are given.

  – [103] A necessary and sufficient criterion to establish input-to-state stability (ISS) of nonlinear dynamical systems, the dynamics of which are periodic with respect to certain state variables and which possess multiple invariant solutions (equilibria, limit cycles, etc.), is provided. Unlike standard Lyapunov approaches, the condition is relaxed and formulated via a sign-indefinite function with sign-definite derivative, and by taking the system's periodicity explicitly into account. The new result is established by using the framework of cell structure introduced in [24] and it complements the methods developed in [3], [4] for periodic systems. The efficiency of the proposed approach is illustrated via the global analysis of a nonlinear pendulum with constant persistent input.
  – In [53] we revisit the problem of stabilizing a triple integrator using a control that depends on the signs of the state variables. For a more general class of linear systems it is shown that the stabilization by sign feedback is possible, depending on some properties of the system's matrix. The conditions for the stability are established by means of linear matrix inequalities. For the triple integrator, the domain of stability is evaluated. Also, the control law is augmented by a linear feedback and the stability properties for this case, checked. The results are illustrated by numerical experiments for a chain of integrators of third order.

- Stabilization:
  – A solution to the problem of global fixed-time output stabilization of a chain of integrators is proposed in [70]. A nonlinear state feedback and a dynamic observer are designed in order to guarantee both fixed-time estimation and fixed-time control. Robustness with respect to exogenous disturbances and measurement noises is established. The performance of the obtained control and estimation algorithms are illustrated by numeric experiments.
  – In [20], the rate of convergence to the origin for a chain of integrators stabilized by homogeneous feedback is accelerated by a supervisory switching of control parameters. The proposed acceleration algorithm ensures a fixed-time convergence for otherwise exponentially or finite-time stable homogeneous closed-loop systems. Bounded disturbances are taken into account. The results are especially useful in the case of exponentially stable systems widespread in the practice. The proposed switching strategy is illustrated by computer simulation.
  – [33] The problem of robust finite-time stabilization of perturbed multi-input linear system by means of generalized relay feedback is considered. A new control design procedure, which combines convex embedding technique with Implicit Lyapunov Function (ILF) method, is developed. The sufficient conditions for both local and global finite-time stabilization are provided. The issues of practical implementation of the obtained implicit relay feedback are discussed. Our theoretical result is supported by numerical simulation for a Buck converter.
  – [100] contributes to the stability analysis for impulsive dynamical systems based on a vector Lyapunov function and its divergence operator. The new method relies on a 2D time domain representation. The result is illustrated for the exponential stability of linear impulsive systems based on LMIs. The obtained results provide some notions of minimum and maximum dwell-time. Some examples illustrate the feasibility of the proposed approach.

  - The Universal Integral Control, introduced in H.K. Khalil, is revisited in [34] by employing mollifiers instead of a high-gain observer for the differentiation of the output signal. The closed loop system is a classical functional differential equation with distributed delays on which standard Lyapunov arguments are applied to study the stability. Low-pass filtering capability of mollifiers is demonstrated for a high amplitude and rapidly oscillating noise. The controller is supported by numerical simulations.
- Synchronization:
  - In [12], we study a robust synchronization problem for multistable systems evolving on manifolds within an Input-to-State Stability (ISS) framework. Based on a recent generalization of the classical ISS theory to multistable systems, a robust synchronization protocol is designed with respect to a compact invariant set of the unperturbed system. The invariant set is assumed to admit a decomposition without cycles, that is, with neither homoclinic nor heteroclinic orbits. Numerical simulation examples illustrate our theoretical results.
  - In [51], [96], motivated by a recent work of R. Brockett Brockett (2013), we study a robust synchronization problem for multistable Brockett oscillators within an Input-to-State Stability (ISS) framework. Based on a recent generalization of the classical ISS theory to multistable systems and its application to the synchronization of multistable systems, a synchronization protocol is designed with respect to compact invariant sets of the unperturbed Brockett oscillator. The invariant sets are assumed to admit a decomposition without cycles (i.e. with neither homoclinic nor heteroclinic orbits). Contrarily to the local analysis of Brockett (2013), the conditions obtained in our work are global and applicable for family of non-identical oscillators. Numerical simulation examples illustrate our theoretical results.

7.5. Non-Linear, Sampled-Data And Time-Delay Systems

- Time-delay systems:
  - The problem of delay estimation for a class of nonlinear time-delay systems is considered in [82]. The theory of non-commutative rings is used to analyze the identifiability. Sliding mode technique is utilized in order to estimate the delay showing the possibility to have a local (or global) delay estimation for periodic (or aperiodic) delay signals.
  - In [14] we give sufficient conditions guaranteeing the observability of singular linear systems with commensurable delays affected by unknown inputs appearing in both the state equation and the output equation. These conditions allow for the reconstruction of the entire state vector using past and actual values of the system output. The obtained conditions coincide with known necessary and sufficient conditions of singular linear systems without delays.
  - [67] presents a finite-time observer for linear time-delay systems. In contrast to many observers, which normally estimate the system state in an asymptotic fashion, this observer estimates the exact system state in predetermined finite time. The finite-time observer proposed is achieved by updating the observer state based on actual and pass data of the observer. Simulation results are also presented to illustrate the convergence behavior of the finite-time observer.
  - The backward observability (BO) of a part of the vector of trajectories of the system state is tackled in [57] for a general class of linear time-delay descriptor systems with unknown inputs. By following a recursive algorithm, we present easy testable sufficient conditions ensuring the BO of descriptor time-delay systems.

  - Motivated by the problem of phase-locking in droop-controlled inverter-based microgrids with delays, in [23], the recently developed theory of input-to-state stability (ISS) for multistable systems is extended to the case of multistable systems with delayed dynamics. Sufficient conditions for ISS of delayed systems are presented using Lyapunov-Razumikhin functions. It is shown that ISS multistable systems are robust with respect to delays in a feedback. The derived theory is applied to two examples. First, the ISS property is established for the model of a nonlinear pendulum and delay-dependent robustness conditions are derived. Second, it is shown that, under certain assumptions, the problem of phase-locking analysis in droop-controlled inverter-based microgrids with delays can be reduced to the stability investigation of the nonlinear pendulum. For this case, corresponding delay-dependent conditions for asymptotic phase-locking are given.
  - Causal and non-causal observability are discussed in [68] for nonlinear time- delay systems. By extending the Lie derivative for time-delay systems in the algebraic framework introduced by Xia et al. (2002), we present a canonical form and give sufficient condition in order to deal with causal and non-causal observations of state and unknown inputs of time-delay systems.
  - [83] presents a finite-time observer for linear time-delay systems with commensurate delay. Unlike the existing observers in the literature which converges asymptotically, the proposed observer provides a finite-time estimation. This is realized by using the well-known sliding mode technique. Simulation results are also presented in order to illustrate the feasibility of the proposed method.
- Sampled-Data systems:
  - [104] presents basic concepts and recent research directions about the stability of sampled-data systems with aperiodic sampling. We focus mainly on the stability problem for systems with arbitrary time-varying sampling intervals which has been addressed in several areas of research in Control Theory. Systems with aperiodic sampling can be seen as time-delay systems, hybrid systems, Input/Output interconnections, discrete-time systems with time-varying parameters, etc. The goal of the article is to provide a structural overview of the progress made on the stability analysis problem. Without being exhaustive, which would be neither possible nor useful, we try to bring together results from diverse communities and present them in a unified manner. For each of the existing approaches, the basic concepts, fundamental results, converse stability theorems (when available), and relations with the other approaches are discussed in detail. Results concerning extensions of Lyapunov and frequency domain methods for systems with aperiodic sampling are recalled, as they allow to derive constructive stability conditions. Furthermore, numerical criteria are presented while indicating the sources of conservatism, the problems that remain open and the possible directions of improvement. At last, some emerging research directions, such as the design of stabilizing sampling sequences, are briefly discussed.
  - In [31] we investigate the stability analysis of nonlinear sampled-data systems, which are affine in the input. We assume that a stabilizing controller is designed using the emulation technique. We intend to provide sufficient stability conditions for the resulting sampled-data system. This allows to find an estimate of the upper bound on the asynchronous sampling intervals, for which stability is ensured. The main idea of the paper is to address the stability problem in a new framework inspired by the dissipativity theory. Furthermore, the result is shown to be constructive. Numerically tractable criteria are derived using linear matrix inequality for polytopic systems and using sum of squares technique for the class of polynomial systems.
  - [76] deals with the sampled-data control problem based on state estimation for linear sampled-data systems. An impulsive system approach is proposed based on a vector Lyapunov function method. Observer-based control design conditions are expressed in terms of LMIs. Some examples illustrate the feasibility of the proposed approach.

7.6. Effective algebraic systems theory

- Algebraic analysis approach:
  - The purpose of [97] is to present a survey on the effective algebraic analysis approach to linear systems theory with applications to control theory and mathematical physics. In particular, we show how the combination of effective methods of computer algebra – based on Gröbner basis techniques over a class of noncommutative polynomial rings of functional operators called Ore algebras – and constructive aspects of module theory and homological algebra enables the characterization of structural properties of linear functional systems. Algorithms are given and a dedicated implementation, called ORE-ALGEBRAICANALYSIS, based on the Mathematica package HOLONOMICFUNCTIONS, is demonstrated.
  - As far as we know, there is no algebraic (polynomial) approach for the study of linear differential time-delay systems in the case of a (sufficiently regular) time-varying delay. Based on the concept of skew polynomial rings developed by Ore in the 30s, the purpose of [73] is to construct the ring of differential time-delay operators as an Ore extension and to analyze its properties. Classical algebraic properties of this ring, such as noetherianity, its homological and Krull dimensions and the existence of Gröbner bases, are characterized in terms of the time-varying delay function. In conclusion, the algebraic analysis approach to linear systems theory allows us to study linear differential time-varying delay systems (e.g. existence of autonomous elements, controllability, parametrizability, flatness, behavioral approach) through methods coming from module theory, homological algebra and constructive algebra.
  - Within the algebraic analysis approach to linear systems theory, in [98], we investigate the equivalence problem of linear functional systems, i.e., the problem of characterizing when all the solutions of two linear functional systems are in a one-to-one correspondence. To do that, we first provide a new characterization of isomorphic finitely presented modules in terms of inflations of their presentation matrices. We then prove several isomorphisms which are consequences of the unimodular completion problem. We then use these isomorphisms to complete and refine existing results concerning Serre's reduction problem. Finally, different consequences of these results are given. All the results obtained are algorithmic for rings for which Gröbner basis techniques exist and the computations can be performed by the Maple packages OREMODULES and OREMORPHISMS.
  - In [99], we study algorithmic aspects of the algebra of linear ordinary integro-differential operators with polynomial coefficients. Even though this algebra is not Noetherian and has zero divisors, Bavula recently proved that it is coherent, which allows one to develop an algebraic systems theory over this algebra. For an algorithmic approach to linear systems of integro-differential equations with boundary conditions, computing the kernel of matrices with entries in this algebra is a fundamental task. As a first step, we have to find annihilators of integro-differential operators, which, in turn, is related to the computation of polynomial solutions of such operators. For a class of linear operators including integro-differential operators, we present an algorithmic approach for computing polynomial solutions and the index. A generating set for right annihilators can be constructed in terms of such polynomial solutions. For initial value problems, an involution of the algebra of integro-differential operators then allows us to compute left annihilators, which can be interpreted as compatibility conditions of integro-differential equations with boundary conditions.
  - Recent progress in computer algebra has opened new opportunities for the parameter estimation problem in nonlinear control theory, by means of integro-differential input-output equations. In [102], we recall the origin of integro-differential equations. We present new opportunities in nonlinear control theory. Finally, we review related recent theoretical approaches on integro-differential algebras, illustrating what an integro-differential elimination method might be and what benefits the parameter estimation problem would gain from

it.

- Computational real algebraic geometric approach:
  - In [74], we present a symbolic-numeric method for solving the H_∞ loop-shaping design problem for low order single-input single-output systems with parameters. Due to the system parameters, no purely numerical algorithm can indeed solve the problem. Using Gröbner basis techniques and the Rational Univariate Representation of zero-dimensional algebraic varieties, we first give a parametrization of all the solutions of the two Algebraic Riccati Equations associated with the H_∞ control problem. Then, following some works on the spectral factorization problem, a certified symbolic-numeric algorithm is obtained for the computation of the positive definite solutions of these two Algebraic Riccati Equations. Finally, we present a certified symbolic-numeric algorithm which solves the H_∞ loop-shaping design problem for the above class of systems.
  - In [58], the asymptotic stability of linear differential systems with commensurate delays is studied. A classical approach for checking that all the roots of the corresponding quasipolynomial have negative real parts consists in computing the set of critical zeros of the quasipolynomial, i.e., the roots (and the corresponding delays) of the quasipolynomial that lie on the imaginary axis, and then analyzing the variation of these roots with respect to the variation of the delay. Based on solving algebraic systems techniques, a certified and efficient symbolic-numeric algorithm for computing the set of critical roots of a quasipolynomial is proposed. Moreover, using recent algorithmic results developed by the computer algebra community, we present an efficient algorithm for the computation of Puiseux series at a critical zero which allows us to finely analyze the stability of the system with respect to the variation of the delay.
  - In [59], we present new computer algebra based methods for testing the structural stability of n-D discrete linear systems (with $n \geq 2$). More precisely, we show that the standard characterization of the structural stability of a multivariate rational transfer function (namely, the denominator of the transfer function does not have solutions in the unit polydisc of $\mathbb{C}^n$) is equivalent to fact that a certain system of polynomials does not have real solutions. We then use state-of-the-art algorithms of the computer algebra community to check this last condition, and thus the structural stability of multidimensional systems.

7.7. Applications

- A fault detection method for an automatic detection of spawning in oysters [13]:

  Using measurements of valve activity (i.e. the distance between the two valves) in populations of bivalves under natural environmental condition (16 oysters in the Bay of Arcachon, France, in 2007, 2013 and 2014), an algorithm for an automatic detection of the spawning period of oysters is proposed in this paper. Spawning observations are important in aquaculture and biological studies, and until now, such a detection is done through visual analysis by an expert. The algorithm is based on the fault detection approach and it works through the estimation of velocity of valve movement activity, that can be obtained by calculating the time derivative of the valve distance. A summarized description of the methods used for the derivative estimation is provided, followed by the associated signal processing and decision making algorithm to determine spawning from the velocity signal. A protection from false spawning detection is also considered by analyzing the simultaneity in spawning. Through this study, it is shown that spawning in a population of oysters living in their natural habitat (i.e. in the sea) can be automatically detected without any human expertise saving time and resources. The fault detection method presented in the paper can also be used to detect complex oscillatory behavior which is of interest to control engineering community.
- Robust synchronization of genetic oscillators [52]:

Cell division introduces discontinuities in the dynamics of genetic oscillators (circadian clocks, synthetic oscillators, etc.) causing phase drift. This paper considers the problem of phase synchronization for a population of genetic oscillators that undergoes cell division and with a common entraining input in the population. Inspired by stochastic simulation, this paper proposes analytical conditions that guarantee phase synchronization. This analytical conditions are derived based on Phase Response Curve (PRC) model of an oscillator (the first order reduced model obtained for the linearized system and inputs with sufficiently small amplitude). Cell division introduces state resetting in the model (or phase resetting in the case of phase model), placing it in the class of hybrid systems. It is shown through numerical experiments for a motivating example that without common entraining input in all oscillators, the cell division acts as a disturbance causing phase drift, while the presence of entrainment guarantees boundedness of synchronization phase errors in the population. Theoretical developments proposed in the paper are demonstrated through numerical simulations for two different genetic oscillator models (Goodwin oscillator and Van der Pol oscillator).

- Modeling pointing tasks in mouse-based human-computer interactions [54]:

  Pointing is a basic gesture performed by any user during human-computer interaction. It consists in covering a distance to select a target via the cursor in a graphical user interface (e.g. a computer mouse movement to select a menu element). In this work, a dynamic model is proposed to describe the cursor motion during the pointing task. The model design is based on experimental data for pointing with a mouse. The obtained model has switched dynamics, which corresponds well to the state of the art accepted in the human-computer interaction community. The conditions of the model stability are established. The presented model can be further used for the improvement of user performance during pointing tasks.

- Modeling and control of turbulent flows [64]:

  The model-based closed-loop control of a separated flow can be studied based on the model described by Navier-Stokes equation. However, such a model still rises difficult issues for control practice. An alternative bilinear and delayed model has been developed tested on the experiments allowing its identification. The identification technique combines least-square technique with a Mesh Adaptive Direct Search (MADS) algorithm.

- Practical design considerations for successful industrial application of model-based fault detection techniques to aircraft systems [47]:

  This paper discusses some key factors which may arise for successful application of model-based Fault Detection(FD) techniques to aircraft systems. The paper reports on the results and the lessons learned during flight V & V(Validation & Verification) activities, implementation in the A380 Flight Control Computer(FCC) and A380 flight tests at Airbus(Toulouse, France).The paper does not focus on new theoretical materials, but rather on a number of practical design considerations to provide viable technological solutions and mechanization schemes. The selected case studies are taken from past and on-going research actions between Airbus and the University of Bordeaux (France). One of the presented solutions has received final certification on new generation Airbus A350 aircraft and is flying (first commercial flight: January 15,2015)

- Finite-time obstacle avoidance for unicycle-like robot [26]:

  The problem of avoiding obstacles while navigating within an environment for a Unicycle-like Wheeled Mobile Robot (WMR) is of prime importance in robotics; the aim of this work is to solve such a problem proposing a perturbed version of the standard kinematic model able to compensate for the neglected dynamics of the robot. The disturbances are considered additive on the inputs and the solution is based on the supervisory control framework, finite-time stability and a robust multi-output regulation. The effectiveness of the solution is proved, supported by experiments and finally compared with the Dynamic Window Approach (DWA) to show how the proposed method can perform better than standard methods.

- Almost global attractivity of a synchronous generator connected to an infinite bus [56]:

The problem of deriving verifiable conditions for stability of the equilibria of a realistic model of a synchronous generator with constant field current connected to an infinite bus is studied in the paper. Necessary and sufficient conditions for existence and uniqueness of equilibrium points are provided. Furthermore, sufficient conditions for almost global attractivity are given. To carry out this analysis a new Lyapunov–like function is proposed to establish convergence of bounded trajectories, while the latter is proven using the powerful theoretical framework of cell structures pioneered by Leonov and Noldus.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Contract with Neotrope (Tourcoing, France), Technologies & Augmented Human UX. Subject: De-correlation of GSR measurements with acceleration, from March 2016 to September 2016, D. Efimov, R. Ushirobira.

8.2. Bilateral Grants with Industry

Project of Autonomous control of clinic table with La Maison Attentive, 2016.

8.3. Bilateral Grants with Industry

Collaboration with Safran Electronics & Defense (Massy-Palaiseau) in the framework of the CIFRE PhD thesis of Guillaume Rance on robust stabilization of gyrostabilized platforms (2014-2018).

9. Partnerships and Cooperations

9.1. Regional Initiatives

- Project ARCIR ESTIREZ “Estimation distribuée de systèmes dynamiques en réseaux”, coordinators: D. Efimov, M. Petreczky, 2013-2017.
- CPER DATA 2016-2020 (involved in two projects: “FIT” related to the wireless robots and sensors network and “DATA”, related to platform). FIT includes our robotic activity and DATA corresponds to our computation need in fluid mechanics as well as possible security issues in the ControlHub development platform.
- ELSAT20202 (Ecomobilité, Logistique, Sécurité, Adaptabilité dans les Transports) is a Regional consortium gathering aeronautics (ONERA), micro/nano technologies (IEMN), control sciences (Non-A) and fluid mechanics (LAMIH, LML) and working on technologies and methods for the active control of separated flows.

9.2. National Initiatives

- ANR project Finite4SoS (Finite time control and estimation for Systems of Systems), coordinator: W. Perruquetti, 2015-2020.
- ANR project WaQMoS (Coastal waters quality surveillance using bivalve mollusk-based sensors), coordinator: D. Efimov, 2015-2019.
- ANR project TurboTouch (High-performance touch interactions), coordinator: G. Casiez (MJOLNIR team, Inria), 2014-2018.
- ANR project ROCC-SYS (Robust Control of Cyber-Physical Systems), coordinator: L. Hetel (CNRS, EC de Lille), 2013-2018.

- ANR project MSDOS (Multidimentional System: Digression on Stability), coordinator: Nima Yeganefar (Poitiers University), 2014-2018.
- We are also involved in several technical groups of the GDR MACS (CNRS, "Modélisation, Analyse de Conduite des Systèmes dynamiques", see http://www.univ-valenciennes.fr/GDR-MACS), in particular: Technical Groups "Identification", "Time Delay Systems", "Hybrid Systems", "Complex Systems, Biological Systems and Automatic Control," and "Control in Electrical Engineering".
- Model-free control: collaborations with the startup ALIEN SAS (created by C. Join and M. Fliess).

9.3. European Initiatives

9.3.1. Collaborations with Major European Organizations

Partner 1: KULeuven, labo 1 (Belgium)

Supervisor: W. Michiels

Partner 2: TU/Eindhoven, labo 1 (The Netherlands)

Supervisor: H. Nijmeijer

Partner 3: Centrale Lille, labo 1 (France)

Supervisor: J.-P. Richard

H2020 project UCoCoS ("Understanding and Controlling of Complex Systems", 2016-2020) is a European Joint Doctorate aiming at creating a framework for complex systems, and at defining a common language, common methods, tools and software for the complexity scientist. It strongly relies on a control theory point of view. Six ESR (early stage researchers) perform a cutting-edge project, strongly relying on the complementary expertise of the 3 academic beneficiaries and benefiting from training by 4 non-academic partners from different sectors. ESR1: Analytical and numerical bifurcation analysis of delay-coupled systems; ESR2: Estimation in complex systems; ESR3: Grip on partial synchronization in delay-coupled networks; ESR4: Reduced modelling of large-scale networks ; ESR5: Network design for decentralized control ; ESR6: Networks with event triggered computing. Non-A is firstly invested on ESR 2 (Haik Silm), 4 (Quentin Voortman), 5 (Deesh Dileep), 6 (Jijju Thomas).

9.4. International Initiatives

9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

HoTSMoCE "Homogeneity Tools for Sliding Mode Control and Estimation", project with UNAM (Mexico), supervisor: D. Efimov, 2015-2018.

9.4.2. Inria International Partners

9.4.2.1. Declared Inria International Partners

Arie Levant, Tel Aviv University, Israel (Invited Professor, 4 months, 2015-2016).

9.4.2.2. Informal International Partners

- Emilia Fridman, Tel Aviv University, Israel
- Leonid Fridman, UNAM, Mexico
- Jaime Moreno, UNAM, Mexico
- Johannes Schiffer, Leeds University, UK
- ITMO University, Saint-Petersburg, Russia
- Eva Zerz, Aachen University, Germany

9.4.3. Participation in Other International Programs

- “Robust and Reliable Control of Aerial Syste”’, Beihang University, China, 2016, in charge: G. Zheng
- PHC Amadeus “Computer Algebra and Functional Equations”, 2016-2017, with the University of Limoges (XLIM) and the University of Linz (Austria).

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Leonid Fridman, UNAM, Mexico, 10/07/2016-22/07/2016, “Stability analysis of a sliding-mode control algorithm of second order with time delays”.
- Emilia Fridman, Tel Aviv University, Israel, 27/06/2016-11/07/2016, “Design of interval observers for distributed-parameters systems”.
- Jaime Alberto Moreno Pérez, UNAM, Mexico, 27/06/2016-08/07/2016, “Recursive design of Lyapunov functions for finite-time stable systems”.
- Tonametl Sanchez Ramirez, UNAM, Mexico, 24/10/2016-18/11/2016, “Homogeneity for discrete-time systems”.
- Juan Gustavo Rueda Escobedo, UNAM, Mexico, 24/10/2016-18/11/2016, “Finite-time and fixed-time identification of parameters”.
- Konstantin Zimenko, ITMO, Russia, 26/09/2016-28/10/2016, “Delay independent stabilization via implicit Lyapunov function approach”.
- Damiano Rotondo, NTNU, Norway, 17/10/2016-21/10/2016, “Fault detection for LPV systems using interval observers”.

9.5.1.1. Internships

- Paul Lesur, “Robust control of blimp”, 05-07/2016 , supervisor: G. Zheng
- Baihui Du, “Robust control of fast dynamical systems”, 05-07/2016, supervisor: G. Zheng

9.5.2. Visits to International Teams

G. Zheng visited Beihang University (China) for two weeks in July 2016.

9.5.2.1. Explorer programme

COSY (under evaluation) Real-time Control of Synthetic microbial communities. While some precursory work has appeared in recent years, the control of microbial communities remains largely unexplored. This proposal aims at exploiting the potential of state-of-art biological modelling, control techniques, synthetic biology and experimental equipment to achieve a paradigm shift in control of microbial communities. Lead by E. Cinquemani as a collaboration of 4 Inria teams IBIS, BIOCORE, COMMANDS, Non-A), the Inria Exploratory Action INBIO and external partners BIOP (CNRS), MaIAge (INRA), and YoukLAB (TU Delft).

9.5.2.2. Research Stays Abroad

G. Zheng held a visiting professor position in Nanjing University of Science and Technology (China) for two months stay in August 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- W. Perruquetti is the chairman of the IFAC Technical Committee “Social Impact of Automation”, International Federation of Automatic Control, TC 9.2, and a member of the IFAC Technical Committees “Nonlinear Control Systems”, TC 2.3, and “Discrete Event and Hybrid Systems”, TC 1.3.
- A. Quadrat is a member of the IFAC Technical Committee “Linear Control Systems”, International Federation of Automatic Control, TC2.2
- J.-P. Richard is a member of the IFAC Technical Committee “Linear Control Systems”, International Federation of Automatic Control, TC2.2
- G. Zheng is a member of the IFAC Technical Committee “Social Impact of Automation”, International Federation of Automatic Control, TC9.2
- G. Zheng is co-chair of the working group “Commande et pilotage en environnement incertain” of GRAISYHM

10.1.1.2. Member of the Organizing Committees

C. Jamroz, A. Quadrat, J.-P. Richard and G. Zheng were members of the organizing committee of “JAMACS’16 : Journées Automatique du GDR MACS 2016”, Villeneuve d’Ascq, 15-16/11/2016.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

W. Perruquetti is a member of the steering committee of IFAC CPHS’16 (7-9 December 2016, Florianopolis, Brazil), an IPC member of IEEE VSS’16 (1-4 June 2016, Nanjing, Jiangsu, China) and of IFAC HMS’16 (August 30-September 2 2016, Kyoto, Japan) and an Associate Editor of the 20th IFAC World Congress (10-14 July 2017, Toulouse, France).

W. Perruquetti and J.-P. Richard are members of the Advisory panel (NOC) of 20th IFAC World Congress, Toulouse, France, 10-14 July 2017.

10.1.2.2. Reviewer

The members of NON-A team are reviewers and contributors of all top-ranked conferences in the field of automatic control (IEEE Conference on Decision and Control, IFAC World Congress, European Control Conference, American Control Conference, etc.).

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- A. Polyakov: International Journal of Robust and Nonlinear Control
- A. Polyakov: Journal of Optimization Theory and Applications (JOTA)
- A. Polyakov: Automation and Remote Control
- A. Quadrat: Multidimensional Systems and Signal Processing (MSSP)

10.1.3.2. Reviewer - Reviewing Activities

The members of NON-A team are reviewers of all top-ranked journals in the field of automatic control (IEEE Transactions on Automatic Control, Automatica, SIAM Journal of Control and Optimization, International Journal of Robust and Nonlinear Control, etc.).

10.1.4. Invited Talks

D. Efimov gave a plenary talk at the conference “JAMACS’16 : Journées Automatique du GDR MACS 2016”, Villeneuve d’Ascq, France, 15-16/11/2016.

10.1.5. Leadership within the Scientific Community

The NON-A team is the leader in the field of non-asymptotic control and estimation using homogeneity framework.

Moreover, the NON-A team is also leader in algebraic systems theory. In particular, two invited sessions "Algebraic methods and symbolic-numeric computation in systems theory" were organized at the 22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS 2016), University of Minnesota, USA, 12-15/07/2016. Moreover, a mini-workshop "New trends on multidimensional systems and their applications in control theory and signal processing" was organized at Centre International de Rencontres Mathématiques (CIRM), Luminy, France, 03-07/10/2016.

10.1.6. Scientific Expertise

A. Quadrat was a member of the "jury d'admission des concours Inria CR2 & CR1". He was also a member of the "Commission des Emplois de Recherche", Inria Lille, and a member of the "Authorité de déchiffrement" for the local elections.

10.1.7. Research Administration

- W. Perruquetti is Vice-deputy of INS2I CNRS.
- J.-P. Richard is an Expert for the French Ministry of Research, MENESR/MEIRIES.
- R. Ushirobira was a Member (nominated) of the "Comité du Centre" of Inria Lille (Dec. 2013 - Sept. 2016) and a Member of the "Commission de Développement Technologique" (CDT).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence: D. Efimov, Laboratory works in automatics, 20 h, EC de Lille.

Licence: D. Efimov, Practical works in automatics, 28 h, ISEN, Lille.

Licence: D. Efimov, Laboratory works in discrete systems, 20 h, ENSAM, Lille.

Licence: R. Ushirobira, Travaux Pratiques en Automatique, 8 h, U. Lille 1, France.

Licence: R. Ushirobira, Travaux Pratiques en Automatique, 24 h, EC-Lille, France.

Licence: R. Ushirobira, Travaux dirigés/Travaux Pratiques en Automatique, 12 h+9 h, U. Lille 1.

Master: D. Efimov, Analysis of dynamical systems, 24 h, U. Lille 1.

Master: R. Ushirobira, Travaux Pratiques en Automatique, 32 h, U. Lille 1.

Master: G. Zheng, Robotic, 20 h, U. Lille 1.

Master: G. Zheng, Automatic control, 24 h, U. Lille 1.

10.2.2. Supervision

PhD: Hafiz Ahmed, "Modeling and synchronization of biological rhythms: from cells to oyster behavior", 2013-2016, supervisors: D. Efimov, R. Ushirobira and D. Tran

PhD: Zilong Shao, "Oscilatory control of robot manipulator", EC Lille, 2012-2016, supervisors: D. Efimov, W. Perruquetti, G. Zheng

PhD in progress: Haik Jan Davtjan, "Estimation in complex systems", EC Lille, 2016, UCoCoS EU project, supervisors: D. Efimov, J.-P. Richard

PhD in progress: Maxime Feingesicht, "Dynamic Observers for Control of Separeted Flows", Ecole Centrale de Lille, 2015, supervisors: J.-P. Richard, F. Kerherve, A. Polyakov

PhD in progress: Nadhynee Martinez Fonseca, "Non-asymptotic control and estimation problems in robotic system designed for manipulation of micro-organisms", National Polytechnic Institute of Mexico, 2015-now, supervisors: I. Chairez-Oria, A. Polyakov

PhD in progress: Tatiana Kharkovskaya, "Interval Observers for Distributed Parametr Systems", ITMO University-EC Lille, 2015, supervisors: D. Efimov, J.-P. Richard and A. Kremlev

PhD in progress: Langueh Désiré Kokou, "Inversion a gauche, singularités d'inversion, immersion et formes normales pour les systèmes dynamiques", 2015, supervisors: T. Floquet, G. Zheng

PhD in progress: Gabriele Perozzi, “Save exploration of aerodynamic field by microdron”, Onera-Region, 2015, supervisors: D. Efimov, J.-M. Biannic and L. Planckaert

PhD in progress: Francisco Lopez-Ramirez, “Control and estimation via implicit homogeneous Lyapunov function”, Inria, 2015, supervisors: D. Efimov, W. Perruquetti and A. Polyakov

PhD in progress: Guillaume Rance, “Asservissement paramétrique de systèmes flexibles à retard et application aux viseurs”, CIFRE Safran Electronics & Defense, 2014, supervisors: A. Quadrat, A. Quadrat, H. Mounier

PhD in progress: Haik-Jan Silm, “Estimation in complex systems”, 2016, supervisors: D. Efimov, R. Ushirobira, W. Michels, J.-P. Richard

PhD in progress: Yue Wang, “Development of a blimp robot for indoor operation”, EC Lille, 2016, supervisors: D. Efimov, W. Perruquetti, G. Zheng

Master: Boussad Abci, EC Lille, 2015-2016, supervisors: D. Efimov, J.-P. Richard

Master: Rabehi Djahid, EC Lille, 2015-2016, supervisors: D. Efimov, J.-P. Richard

10.2.3. Juries

- A. Quadrat was an Examiner Member of the PhD Thesis of Mohamed Belhocine, “Modélisation et analyse structurelle du fonctionnement dynamique des systèmes électriques”, ENS Cachan. He was also a Member of Recruiting Committee for a MCF CNU 26-27 position at the University of Limoges.
- J.-P. Richard was an Examiner Member of the PhD Thesis of Arvo Kaldmae (Estonia), “Design of discrete-time and delayed nonlinear systems”, of Lucien Etienne (Italy), “Elements of observation and estimation for networked control systems”, and of Zilong Shao (Centrale Lille) “Identification and control of position-controlled robot arm in the presence of joint flexibility”.
- R. Ushirobira was a Member of Recruiting Committee for a MCF CNU 61 position at CNAM (Paris) and for a MCF CNU 61 position at ENSAE (Cergy).

10.3. Popularization

- Mediation: Scientific baccalaureate students. Meeting on the Inria platform EuraTechnologies (23/03/2016, Lille): “SN Lille Académie : À la découverte des sciences numériques !”.
- Mediation: BeyondLab community. A co-working night event within BeyondLabon using “Living sensor” for water quality monitoring, featuring our PhD student Hafiz Ahmed (Lille, 16/03/2016).
- Contribution to the prospective report “Systems & Control for the Future of Humanity” coordinated by F. Lamnabhi-Lagarrigue, Research Agenda Task Force (to appear, special issue of Annual Reviews in Control to be distributed during the IFAC 2017).
- The book “Mathématiques pour l’ingénieur” (2009, ISBN : 978-9973-0-0852-7 (ATAN, 385 pages)) has been downloaded more than 65000 times.

11. Bibliography

Major publications by the team in recent years

[1] H. AHMED, R. USHIROBIRA, D. EFIMOV, D. TRAN, M. SOW, L. PAYTON, J.-C. MASSABUAU.*A fault detection method for an automatic detection of spawning in oysters*, in "IEEE Transactions on Control Systems Technology", December 2015, p. 1–8, https://hal.inria.fr/hal-01185118.

[2] F. J. BEJARANO, G. ZHENG.*Observability of linear systems with commensurate delays and unknown inputs*, in "Automatica", April 2014, vol. 50, n^{o} 8, p. 2077-2083 [*DOI :* 10.1016/J.AUTOMATICA.2014.05.032], https://hal.inria.fr/hal-00967944.

[3] L. BELKOURA, T. FLOQUET, K. IBN TAARIT, W. PERRUQUETTI, Y. ORLOV.*Estimation problems for a class of impulsive systems*, in "International Journal of Robust and Nonlinear Control", 2011, vol. 21, n° 10, p. 1066-1079, https://hal.archives-ouvertes.fr/inria-00504941.

[4] E. BERNUAU, A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*Verification of ISS, iISS and IOSS properties applying weighted homogeneity*, in "Systems & Control Letters", December 2013, vol. 62, n° 12, p. 1159-1167, https://hal.inria.fr/hal-00877148.

[5] D. EFIMOV, W. PERRUQUETTI, J.-P. RICHARD.*Interval estimation for uncertain systems with time-varying delays*, in "International Journal of Control", 2013, vol. 86, n° 10, p. 1777-1787, https://hal.inria.fr/hal-00813314.

[6] D. EFIMOV, W. PERRUQUETTI, J.-P. RICHARD.*Development Of Homogeneity Concept For Time-Delay Systems*, in "SIAM Journal on Control and Optimization", December 2014, https://hal.inria.fr/hal-00956878.

[7] M. FLIESS, C. JOIN.*Model-free control*, in "International Journal of Control", December 2013, vol. 86, n° 12, p. 2228-2252 [*DOI :* 10.1080/00207179.2013.810345], https://hal-polytechnique.archives-ouvertes.fr/hal-00828135.

[8] A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*Finite-time and fixed-time stabilization: Implicit Lyapunov function approach*, in "Automatica", 2015, vol. 51, p. 332 - 340 [*DOI :* 10.1016/J.AUTOMATICA.2014.10.082], https://hal.inria.fr/hal-01098099.

[9] A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI, J.-P. RICHARD.*Implicit Lyapunov-Krasovski Functionals for Stability Analysis and Control Design of Time-Delay Systems*, in "IEEE Transactions on Automatic Control", 2015, vol. 60, n° 12, p. 3344 - 3349, https://hal.inria.fr/hal-01160061.

[10] A. POLYAKOV.*Nonlinear feedback design for fixed-time stabilization of Linear Control Systems*, in "IEEE Transactions on Automatic Control", August 2012, vol. 57, n° 8, p. 2106-2110, https://hal.inria.fr/hal-00757561.

Publications of the year

Articles in International Peer-Reviewed Journal

[11] H. ABOUAÏSSA, M. FLIESS, C. JOIN.*On ramp metering: Towards a better understanding of ALINEA via model-free control*, in "International Journal of Control", 2017 [*DOI :* 10.1080/00207179.2016.1193223], https://hal-polytechnique.archives-ouvertes.fr/hal-01326514.

[12] H. AHMED, R. USHIROBIRA, D. EFIMOV, W. PERRUQUETTI.*Robust synchronization for multistable systems*, in "IEEE Transactions on Automatic Control", May 2016, p. 1–6, https://hal.inria.fr/hal-01185112.

[13] H. AHMED, R. USHIROBIRA, D. EFIMOV, D. TRAN, M. SOW, L. PAYTON, J.-C. MASSABUAU.*A fault detection method for an automatic detection of spawning in oysters*, in "IEEE Transactions on Control Systems Technology", March 2016, vol. 24, n° 3, p. 1140–1147, https://hal.inria.fr/hal-01185118.

[14] F. J. BEJARANO, G. ZHENG.*Observability of singular time-delay systems with unknown inputs*, in "Systems and Control Letters", March 2016, vol. 89, p. 55–60 [*DOI :* 10.1016/J.SYSCONLE.2015.12.002], https://hal.inria.fr/hal-01410648.

[15] D. BENMERZOUK, J.-P. BARBOT.*Symmetries impact in chaotification of piecewise smooth systems*, in "Nonlinear Dynamics and Systems Theory", 2016, vol. 16, n° 2, p. 149-164, https://hal.archives-ouvertes.fr/hal-01322942.

[16] M. DEFOORT, G. DEMESURE, Z. ZUO, A. POLYAKOV, M. DJEMAI.*Fixed-time stabilization and consensus of nonholonomic systems*, in "IET Control Theory and Applications", September 2016 [*DOI :* 10.1049/IET-CTA.2016.0094], https://hal.inria.fr/hal-01405104.

[17] D. EFIMOV, C. EDWARDS, A. ZOLGHADRI.*Enhancement of adaptive observer robustness applying sliding mode techniques*, in "Automatica", December 2016, https://hal.inria.fr/hal-01315009.

[18] D. EFIMOV, E. FRIDMAN, A. POLYAKOV, W. PERRUQUETTI, J.-P. RICHARD.*Linear interval observers under delayed measurements and delay-dependent positivity*, in "Automatica", December 2016, https://hal.inria.fr/hal-01312981.

[19] D. EFIMOV, E. FRIDMAN, A. POLYAKOV, W. PERRUQUETTI, J.-P. RICHARD.*On Design of Interval Observers with Sampled Measurement*, in "Systems and Control Letters", December 2016, https://hal.inria.fr/hal-01353919.

[20] D. EFIMOV, A. LEVANT, A. POLYAKOV, W. PERRUQUETTI.*Supervisory Acceleration of Convergence for Homogeneous Systems*, in "International Journal of Control", February 2017 [*DOI :* 10.1080/00207179.2016.1269949], https://hal.inria.fr/hal-01411088.

[21] D. EFIMOV, W. PERRUQUETTI.*On Conditions of Oscillations and Multi-Homogeneity*, in "Mathematics of Control, Signals, and Systems", February 2016, vol. 28, https://hal.inria.fr/hal-01223154.

[22] D. EFIMOV, T. RAÏSSI.*Design of interval observers for uncertain dynamical systems*, in "Automation and Remote Control / Avtomatika i Telemekhanika", 2016, vol. 77, n° 2, 191-225 [*DOI :* 10.1134/S0005117916020016], https://hal.inria.fr/hal-01276439.

[23] D. EFIMOV, J. SCHIFFER, R. ORTEGA.*Robustness of delayed multistable systems with application to droop-controlled inverter-based microgrids*, in "International Journal of Control", 2016, vol. 89, n° 5, p. 909-918, https://hal.inria.fr/hal-01211454.

[24] A. ESTRADA, D. EFIMOV, W. PERRUQUETTI.*Finite-Time Position and Velocity Estimation Adapted to Noisy Biased Acceleration Measurements from Periodic Motion*, in "International Journal of Control", March 2016, https://hal.inria.fr/hal-01295716.

[25] G. GUERRA, D. EFIMOV, G. ZHENG, W. PERRUQUETTI.*Avoiding local minima in the potential field method using input-to-state stability*, in "Control Engineering Practice", 2016, vol. 55, p. 174 - 184 [*DOI :* 10.1016/J.CONENGPRAC.2016.07.008], https://hal.inria.fr/hal-01405349.

[26] M. GUERRA, D. EFIMOV, G. ZHENG, W. PERRUQUETTI.*Finite-Time Obstacle Avoidance for Unicycle-like robot Subject to Additive Input Disturbances*, in "Autonomous Robots", February 2016, https://hal.inria.fr/hal-01232203.

[27] D.-Y. LIU, G. ZHENG, D. BOUTAT, H.-R. LIU.*Non-asymptotic fractional order differentiator for a class of fractional order linear systems*, in "Automatica", 2017, https://hal.archives-ouvertes.fr/hal-01420910.

[28] S. MAALEJ, A. KRUSZEWSKI, L. BELKOURA.*Robust Control for Continuous LPV System with Restricted-Model-Based Control*, in "Circuits, Systems, and Signal Processing", September 2016 [*DOI :* 10.1007/s00034-016-0404-6], https://hal.archives-ouvertes.fr/hal-01410300.

[29] M. MBAYE, M. DIALLO, M. MBOUP.*LU based Beamforming schemes for MIMO systems*, in "IEEE Transactions on Vehicular Technology", 2016, https://hal.inria.fr/hal-01307533.

[30] S. C. OLTEANU, A. AITOUCHE, L. BELKOURA, A. JOUNI.*'Embedded PEM Fuel cell Stack. Nonlinear Observer by means of TS approach*, in "Studies in Informatics and Control", February 2016, vol. 24, n^o 1, p. 61-70, https://hal.archives-ouvertes.fr/hal-01291046.

[31] H. OMRAN, L. HETEL, M. PETRECZKY, J.-P. RICHARD, F. LAMNABHI-LAGARRIGUE.*Stability analysis of some classes of input-affine nonlinear systems with aperiodic sampled-data control*, in "Automatica", 2016, vol. 70, p. 266–274 [*DOI :* 10.1016/j.automatica.2016.02.013], https://hal.archives-ouvertes.fr/hal-01343851.

[32] A. POLYAKOV, D. EFIMOV, E. FRIDMAN, W. PERRUQUETTI.*On Homogeneous Distributed Parameter Systems*, in "IEEE Transactions on Automatic Control", 2016 [*DOI :* 10.1109/TAC.2016.2525925], https://hal.inria.fr/hal-01318134.

[33] A. POLYAKOV, L. HETEL.*Relay Control Design for Robust Stabilization in a Finite-Time*, in "IEEE Transactions on Automatic Control", 2016 [*DOI :* 10.1109/TAC.2016.2591725], https://hal.inria.fr/hal-01347196.

[34] S. RIACHY, D. EFIMOV, M. MBOUP.*Universal integral control: An approach based on mollifiers*, in "IEEE Transactions on Automatic Control", January 2016, 16, https://hal.inria.fr/hal-01145601.

[35] H. RÍOS, D. EFIMOV, L. FRIDMAN, J. MORENO, W. PERRUQUETTI.*Homogeneity based Uniform Stability Analysis for Time-Varying Systems*, in "IEEE Transactions on Automatic Control", April 2016, p. 1–11, https://hal.inria.fr/hal-01162678.

[36] H. RÍOS, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*Homogeneous Time-Varying Systems: Robustness Analysis*, in "IEEE Transactions on Automatic Control", March 2017, n^o 3, https://hal.inria.fr/hal-01284741.

[37] H. RÍOS, M. MERA, D. EFIMOV, A. POLYAKOV.*Robust output-control for uncertain linear systems: Homogeneous differentiator-based observer approach*, in "International Journal of Robust and Nonlinear Control", 2016 [*DOI :* 10.1002/rnc.3643], https://hal.inria.fr/hal-01405383.

[38] R. TAMI, G. ZHENG, D. BOUTAT, D. AUBRY, H. WANG.*Partial observer normal form for nonlinear system*, in "Automatica", 2016, vol. 64, p. 54–62 [*DOI :* 10.1016/j.automatica.2015.10.041], https://hal.inria.fr/hal-01239583.

[39] R. USHIROBIRA, W. PERRUQUETTI, M. MBOUP.*An algebraic continuous time parameter estimation for a sum of sinusoidal waveform signals*, in "International Journal of Adaptive Control and Signal Processing", 2016 [*DOI :* 10.1002/acs.2688], https://hal.inria.fr/hal-01342210.

[40] H. WANG, Y. XUEFEI, Y. TIAN, G. ZHENG, N. CHRISTOV.*Model-Free Based Terminal SMC of Quadrotor Attitude and Position*, in "IEEE Transactions on Aerospace and Electronic Systems", October 2016, https://hal.inria.fr/hal-01410899.

[41] L. YU, G. ZHENG, D. BOUTAT.*Adaptive observer for simultaneous state and parameter estimations for an output depending normal form*, in "Asian Journal of Control", January 2017, vol. 19, n^{o} 1, p. 1-6, https://hal.inria.fr/hal-01410734.

[42] G. ZHENG, F. J. BEJARANO.*Observer design for linear singular time-delay systems*, in "Automatica", 2017, https://hal.archives-ouvertes.fr/hal-01420897.

[43] G. ZHENG, D. EFIMOV, F. J. BEJARANO, W. PERRUQUETTI, H. WANG.*Interval observer for a class of uncertain nonlinear singular systems*, in "Automatica", 2016, vol. 71, p. 159-168, https://hal.inria.fr/hal-01292475.

[44] S. ZHUK, A. POLYAKOV, O. NAKONECHNYI.*Note on Minimax Sliding Mode Control Design for Linear Systems*, in "IEEE Transactions on Automatic Control", 2016 [*DOI :* 10.1109/TAC.2016.2612058], https://hal.inria.fr/hal-01371091.

[45] K. ZIMENKO, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*A note on delay robustness for homogeneous systems with negative degree*, in "Automatica", June 2017, https://hal.inria.fr/hal-01419131.

[46] D. ZIMENKO, A. POLYAKOV, D. EFIMOV, A. KREMLEV.*Feedback sensitivity functions analysis of finite-time stabilizing control system*, in "International Journal of Robust and Nonlinear Control", 2016 [*DOI :* 10.1002/RNC.3693], https://hal.inria.fr/hal-01405766.

[47] A. ZOLGHADRI, J. CIESLAK, D. EFIMOV, G. PHILIPPE, R. DAYRE.*Practical design considerations for successful industrial application of model-based fault detection techniques to aircraft systems*, in "Annual Reviews in Control", December 2016 [*DOI :* 10.1016/J.ARCONTROL.2016.07.003], https://hal.archives-ouvertes.fr/hal-01385577.

Invited Conferences

[48] J. GAMIOCHIPI, M. GHANES, W. AGGOUNE, J. DELEON, J.-P. BARBOT.*Observer design for a class of nonlinear systems under a persistent excitation*, in "10th IFAC Symposium on Nonlinear Control Systems, NOLCOS,", Monterey, United States, August 2016, https://hal.archives-ouvertes.fr/hal-01333821.

[49] K. A. LANGUEH, O. DATCU, J.-P. BARBOT, G. ZHENG, K. BUSAWON.*Observabilty singularities and observer design: dual immersion approach*, in "10th IFAC Symposium on Nonlinear Control Systems, NOLCOS", Monterey, United States, August 2016, https://hal.archives-ouvertes.fr/hal-01333688.

International Conferences with Proceedings

[50] H. ABOUAÏSSA, M. FLIESS, C. JOIN.*On short-term traffic flow forecasting and its reliability*, in "8th IFAC Conference on Manufacturing Modelling, Management & Control (MIM 2016)", Troyes, France, 8th IFAC Conference on Manufacturing Modelling, Management & Control, IFAC, June 2016, https://hal.archives-ouvertes.fr/hal-01275311.

[51] H. AHMED, R. USHIROBIRA, D. EFIMOV.*On the robust synchronization of Brockett oscillators*, in "PSYCO'2016 - 6th IFAC International Workshop on Periodic Control Systems", Eindhoven, Netherlands, June 2016, vol. 49, p. 142 - 147 [*DOI :* 10.1016/J.IFACOL.2016.07.999], https://hal.inria.fr/hal-01380180.

[52] H. AHMED, R. USHIROBIRA, D. EFIMOV.*Robust synchronization of genetic oscillators subjected to cell division and common entrainment*, in "ECC 2016 - 15th annual European Control Conference", Aalborg, Denmark, June 2016, https://hal.inria.fr/hal-01380247.

[53] A. APARICIO, D. EFIMOV, L. FRIDMAN.*ISS-Lyapunov Functions for Output Feedback Sliding Modes*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371277.

[54] S. ARANOVSKIY, R. USHIROBIRA, D. EFIMOV, G. CASIEZ.*Modeling Pointing Tasks in Mouse-Based Human-Computer Interactions*, in "CDC 2016 - 55th IEEE Conference on Decision and Control", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01380318.

[55] O. BARA, M. FLIESS, C. JOIN, J. DAY, S. M. DJOUADI.*Model-free immune therapy: A control approach to acute inflammation*, in "15th European Control Conference, ECC'16", Aalborg, Denmark, June 2016, https://hal-polytechnique.archives-ouvertes.fr/hal-01341060.

[56] N. BARABANOV, J. SCHIFFER, R. ORTEGA, D. EFIMOV.*Almost Global Attractivity of a Synchronous Generator Connected to an Infinite Bus*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371271.

[57] F. J. BEJARANO, G. ZHENG.*Unknown Input Functional Observability of Descriptor Systems with Delays*, in "CDC2016", Las Vegas, United States, IEEE, December 2016, https://hal.inria.fr/hal-01410874.

[58] Y. BOUZIDI, A. POTEAUX, A. QUADRAT.*Computer algebra methods for the stability analysis of differential systems with commensurate time-delays*, in "13th IFAC Workshop on Time Delay Systems", Istanbul, Turkey, June 2016, 6, https://hal.inria.fr/hal-01415282.

[59] Y. BOUZIDI, F. ROUILLIER.*Certified Algorithms for proving the structural stability of two dimensional systems possibly with parameters*, in "MNTS 2016 - 22nd International Symposium on Mathematical Theory of Networks and Systems", Minneapolis, United States, Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems, July 2016, https://hal.inria.fr/hal-01366202.

[60] K. H. DEGUE, D. EFIMOV, A. IGGIDR.*Interval estimation of sequestered infected erythrocytes in malaria patients*, in "European Control Conference (ECC16)", Aalborg, Denmark, June 2016, vol. CDROM, https://hal.inria.fr/hal-01295741.

[61] D. EFIMOV, A. LEVANT, A. POLYAKOV, W. PERRUQUETTI.*On Acceleration of a Class of Asymptotically Stable Homogeneous Systems*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371274.

[62] D. EFIMOV, R. ORTEGA, J. SCHIFFER.*ISS of multistable systems with delays: application to droop-controlled inverter-based microgrids*, in "ACC'15", Chicago, United States, Proceedings of the 2015 American Control Conference ACC'15, July 2016, https://hal-centralesupelec.archives-ouvertes.fr/hal-01262189.

[63] D. EFIMOV, A. POLYAKOV, A. LEVANT, W. PERRUQUETTI.*Discretization of Asymptotically Stable Homogeneous Systems by Explicit and Implicit Euler Methods*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371275.

[64] M. FEINGESICHT, C. RAIBAUDO, A. POLYAKOV, F. KERHERVE, J.-P. RICHARD.*A bilinear input-output model with state-dependent delay for separated flow control*, in "European Control Conference 2016", Aalborg, Denmark, June 2016, https://hal.inria.fr/hal-01298166.

[65] C. JOIN, M. FLIESS, C. VOYANT, F. CHAXEL.*Solar energy production: Short-term forecasting and risk management*, in "8th IFAC Conference on Manufacturing Modelling, Management & Control (MIM 2016)", Troyes, France, June 2016 [*DOI :* 10.1016/J.IFACOL.2016.07.790], https://hal-polytechnique.archives-ouvertes.fr/hal-01272152.

[66] T. KHARKOVSKAYA, D. EFIMOV, A. POLYAKOV, J.-P. RICHARD.*Interval observers for PDEs: approximation approach*, in "Proc. 10th IFAC Symposium on Nonlinear Control Systems (NOLCOS)", Monterey, United States, August 2016, https://hal.inria.fr/hal-01371283.

[67] K. A. LANGUEH, G. ZHENG, T. FLOQUET.*Finite-time observer for linear system with time delay*, in "IEEE Chinese Control Conference", Chengdu, China, IEEE, July 2016, https://hal.inria.fr/hal-01410852.

[68] K. A. LANGUEH, G. ZHENG, T. FLOQUET.*Finite-time unknown input observer for linear time-delay systems*, in "CDC2016", Las Vegas, United States, IEEE, December 2016, https://hal.inria.fr/hal-01410876.

[69] A. LEVANT, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*Stability and Robustness of Homogeneous Differential Inclusions*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371280.

[70] F. LOPEZ-RAMIREZ, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*Fixed-Time Output Stabilization of a Chain of Integrators*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371278.

[71] F. LOPEZ-RAMIREZ, A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*Finite-Time and Fixed-Time Observers Design via Implicit Lyapunov Function*, in "European Control Conference 2016", Aalborg, Denmark, June 2016, https://hal.inria.fr/hal-01298162.

[72] A. POLYAKOV, J.-M. CORON, L. ROSIER.*On Finite-Time Stabilization of Evolution Equations: A Homogeneous Approach*, in "IEEE Conference on Decision and Control 2016", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371089.

[73] A. QUADRAT, R. USHIROBIRA.*Algebraic analysis for the Ore extension ring of differential time-varying delay operators*, in "22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS)", Minneapolis, United States, July 2016, 8, https://hal.inria.fr/hal-01415256.

[74] G. RANCE, Y. BOUZIDI, A. QUADRAT, A. QUADRAT.*A symbolic-numeric method for the parametric $H\infty$ loop-shaping design problem*, in "22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS)", Minneapolis, United States, July 2016, 8, https://hal.inria.fr/hal-01415294.

[75] H. RÍOS, D. EFIMOV, J. MORENO, W. PERRUQUETTI.*Finite-Time Identification Algorithm based on Time-Varying Homogeneity and Lyapunov Approach*, in "Proc. 10th IFAC Symposium on Nonlinear Control Systems (NOLCOS)", Monterey, United States, August 2016, https://hal.inria.fr/hal-01372047.

[76] H. Ríos, L. Hetel, D. Efimov.*Observer-based Control for Linear Sampled-Data Systems: An Impulsive System Approach*, in "Proc. 55th IEEE Conference on Decision and Control (CDC)", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01371272.

[77] Z. Shao, G. Zheng, D. Efimov, W. Perruquetti.*Robust and adaptive control using measurements of higher order derivatives*, in "ECC 2016", Aalborg, Denmark, IFAC, June 2016, https://hal.inria.fr/hal-01410855.

[78] S. Tebbani, M. Titica, C. Join, M. Fliess, D. Dumur.*Model-based versus model-free control designs for improving microalgae growth in a closed photobioreactor: Some preliminary comparisons*, in "The 24th Mediterranean Conference on Control and Automation (MED'16)", Athens, Greece, The 24th Mediterranean Conference on Control and Automation, June 2016, https://hal-polytechnique.archives-ouvertes.fr/hal-01312251.

[79] R. Ushirobira, D. Efimov, G. Casiez, N. Roussel, W. Perruquetti.*A forecasting algorithm for latency compensation in indirect interaction*, in "In proceedings of ECC'16, the 15th annual European Control Conference", Aalborg, Denmark, June 2016, 6, https://hal.inria.fr/hal-01420653.

[80] R. Ushirobira, A. Quadrat.*Algebraic estimation of a biased and noisy continuous signal via orthogonal polynomials*, in "CDC 2016 - 55th IEEE Conference on Decision and Control", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01380320.

[81] G. Zheng, D. Boutat, H. Wang.*Observer design for a class of nonlinear singular systems*, in "CDC2016", Las Vegas, United States, IEEE, December 2016, https://hal.inria.fr/hal-01410882.

[82] G. Zheng, A. Polyakov, A. Levant.*Delay estimation for nonlinear time-delay systems*, in "International Workshop on Variable Structure Systems", Nanjing, China, IFAC, March 2016, https://hal.inria.fr/hal-01410851.

[83] G. Zheng, H. Wang.*A simple finite-time observer for linear time-delay systems*, in "International Workshop on Variable Structure Systems", Nanjing, China, IFAC, March 2016, https://hal.inria.fr/hal-01410836.

[84] K. Zimenko, D. Efimov, A. Polyakov, W. Perruquetti.*Time-delay Robustness Analysis for Systems with Negative Degree of Homogeneity*, in "Proc. 10th IFAC Symposium on Nonlinear Control Systems (NOLCOS)", Monterey, United States, August 2016, https://hal.inria.fr/hal-01371370.

[85] K. Zimenko, A. Polyakov, D. Efimov, A. Kremlev.*Frequency Domain Analysis of Control System Based on Implicit Lyapunov Function**, in "European Control Conference 2016", Aalborg, Denmark, IFAC, June 2016, https://hal.inria.fr/hal-01298160.

Conferences without Proceedings

[86] J.-P. Barbot, A. Levant, M. Livine, D. Lunz.*Discrete Sliding-Mode-Based Differentiators*, in "The 14th International Workshop on Variable Structure Systems (VSS 2016)", Nanjing, Jiangsu, China, June 2016, https://hal.archives-ouvertes.fr/hal-01338297.

[87] K. H. Degue, D. Efimov, J.-P. Richard.*Interval Observers for Linear Impulsive Systems*, in "10th IFAC Symposium on Nonlinear Control Systems (NOLCOS 2016)", Monterey, California, United States, August 2016, https://hal.inria.fr/hal-01356972.

[88] M. DOUBLET, C. JOIN, F. HAMELIN.*Model-free control for unknown delayed systems*, in "3rd Conference on Control and Fault-Tolerant Systems, SysTol 2016", Barcelone, Spain, September 2016, https://hal.archives-ouvertes.fr/hal-01369776.

[89] D. EFIMOV, E. FRIDMAN, A. POLYAKOV, W. PERRUQUETTI, J.-P. RICHARD.*On Design of Sampled-Data Interval Observers*, in "European Control Conference (ECC16)", Aalborg, Denmark, March 2016, https://hal.inria.fr/hal-01295752.

[90] M. GUERRA, C. VÁZQUEZ, D. EFIMOV, G. ZHENG, L. FREIDOVICH, W. PERRUQUETTI.*Interval Differentiators: on-line estimation of differentiation accuracy*, in "European Control Conference (ECC16)", Aalborg, Denmark, June 2016, https://hal.inria.fr/hal-01295745.

[91] Z. KADER, C. FITER, L. HETEL, L. BELKOURA.*Observer-based relay feedback controller design for LTI systems*, in "European Control Conference (ECC 2016)", Aalborg, Denmark, June 2016, https://hal.archives-ouvertes.fr/hal-01415095.

[92] Z. KADER, C. FITER, L. HETEL, L. BELKOURA.*Stabilization of LTI systems by relay feedback with perturbed measurements*, in "American Control Conference (ACC 2016)", Boston, United States, July 2016, https://hal.archives-ouvertes.fr/hal-01415039.

[93] H. RÍOS, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*Robustness of Homogeneous Systems with Respect to Time-Varying Perturbations* *, in "European Control Conference (ECC16)", Aalborg, Denmark, June 2016, https://hal.inria.fr/hal-01295721.

Scientific Books (or Scientific Book chapters)

[94] E. BERNUAU, D. EFIMOV, W. PERRUQUETTI, A. POLYAKOV.*Homogeneity of differential inclusions*, in "Recent Trends in Sliding Mode Control", 2016, https://hal.inria.fr/hal-01405148.

[95] S. ZHUK, A. POLYAKOV.*Minimax observer for sliding mode control design*, in "Recent Trends in Sliding Mode Control", 2016 [*DOI :* 10.1049/PBCE102E_CH2.2], https://hal.inria.fr/hal-01405115.

Research Reports

[96] H. AHMED, R. USHIROBIRA, D. EFIMOV.*Robust global synchronization of Brockett oscillators*, Inria Lille - Nord Europe ; Clemson University, November 2016, https://hal.inria.fr/hal-01391120.

[97] T. CLUZEAU, C. KOUTSCHAN, A. QUADRAT, M. TÕNSO.*Effective algebraic analysis approach to linear systems over Ore algebras*, Inria Lille - Nord Europe ; University of Limoges, France ; RICAM, Austrian Academy of Sciences ; Institute of Cybernetics, Tallinn University of Technology, December 2016, n^{o} RR-8999, 42, https://hal.inria.fr/hal-01413591.

[98] T. CLUZEAU, A. QUADRAT.*Equivalences of linear functional systems*, Inria Lille - Nord Europe ; University of Limoges, France, December 2016, n^{o} RR-9000, 29, https://hal.inria.fr/hal-01413593.

[99] A. QUADRAT, G. REGENSBURGER.*Computing Polynomial Solutions and Annihilators of Integro-Differential Operators with Polynomial Coefficients*, Inria Lille - Nord Europe ; Institute for Algebra, Johannes Kepler University Linz, December 2016, n^{o} RR-9002, 24, https://hal.inria.fr/hal-01413907.

[100] H. Ríos, L. Hetel, D. Efimov.*Stability Analysis for Impulsive Systems: 2D Vector Lyapunov Function Approach*, Inria Lille - Nord Europe, February 2016, https://hal.inria.fr/hal-01280321.

Patents and standards

[101] S. Aranovskiy, G. Casiez, D. Efimov, N. Roussel, R. Ushirobira.*Dispositif à affichage prédictif*, September 2016, n° 1000365810, https://hal.inria.fr/hal-01426069.

Other Publications

[102] F. Boulier, F. Lemaire, M. Rosenkranz, R. Ushirobira, N. Verdière.*On Symbolic Approaches to Integro-Differential Equations*, March 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01367138.

[103] D. Efimov, J. Schiffer, N. Barabanov, R. Ortega.*A relaxed characterization of ISS for periodic systems with multiple invariant sets*, March 2016, working paper or preprint, https://hal.inria.fr/hal-01351139.

[104] L. Hetel, C. Fiter, H. Omran, A. Seuret, E. Fridman, J.-P. Richard, S.-I. Niculescu.*Recent developments on the stability of systems with aperiodic sampling: an overview* ⋆, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01363448.

[105] A. Quadrat.*Towards an effective study of the algebraic parameter estimation problem*, November 2016, working paper or preprint, https://hal.inria.fr/hal-01415300.

Team RAPSODI

Reliable numerical approximations of dissipative systems

Inria teams are typically groups of researchers working on the definition of a common project, and objectives, with the goal to arrive at the creation of a project-team. Such project-teams may include other partners (universities or research institutions).

RESEARCH CENTER
Lille - Nord Europe

THEME
Numerical schemes and simulations

Table of contents

Team RAPSODI

Creation of the Team: 2015 August 01

Keywords:

Computer Science and Digital Science:

6.1. - Mathematical Modeling
6.1.1. - Continuous Modeling (PDE, ODE)
6.1.3. - Discrete Modeling (multi-agent, people centered)
6.1.4. - Multiscale modeling
6.2. - Scientific Computing, Numerical Analysis & Optimization
6.2.1. - Numerical analysis of PDE and ODE
6.2.5. - Numerical Linear Algebra
6.2.8. - Computational geometry and meshes

Other Research Topics and Application Domains:

3.3.1. - Earth and subsoil
3.3.4. - Atmosphere
3.4.2. - Industrial risks and waste
4. - Energy
5. - Industry of the future
5.2.4. - Aerospace
5.4. - Microelectronics

1. Members

Research Scientist

Clément Cancès [Inria, Researcher, HDR]

Faculty Members

Claire Chainais-Hillairet [Team leader, Univ. Lille I, Professor, HDR]
Caterina Calgaro [Univ. Lille I, Associate Professor]
Pierre-Louis Colin [Univ. Lille I, Associate Professor, until Aug 2016]
Emmanuel Creusé [Univ. Lille I, Professor, HDR]
Ingrid Lacroix Violet [Univ. Lille I, Associate Professor]
Benoît Merlet [[Univ. Lille I, Associate Professor]
Thomas Rey [Univ. Lille I, Associate Professor]

PhD Students

Ahmed Ait Hammou Oulhaj [Univ. Lille I]
Claire Lecerf [Univ. Lille I]
Antoine Zurek [Univ. Lille I, from Apr 2016]

Post-Doctoral Fellow

Flore Nabet [Inria, until Aug 2016]

Administrative Assistant

Corinne Jamroz [Inria]

2. Overall Objectives

2.1. Overall Objectives

Together with the diffusion of scientific computing, there has been a recent and impressive increase of the demand for numerical methods. The problems to be addressed are everyday more complex and require specific numerical algorithms. The quality of the results has to be accurately assessed, so that in-silico experiments results can be trusted. Nowadays, producing such reliable numerical results goes way beyond the abilities of isolated researchers, and must be carried out by structured teams.

The topics addressed by the RAPSODI team belong to the broad theme of numerical methods for the approximation of solutions of systems of partial differential equations (PDEs). Besides standard convergence properties, a good numerical method for approximating a physical problem has to satisfy at least the following three criteria:

1. preservation at the discrete level of some crucial features of the solution, such as positivity of solutions, conservation of prescribed quantities (e.g., mass, the decay of physically motivated entropies, etc.);
2. provide accurate numerical approximations at a reasonable computational cost (and ultimately maximize the accuracy at a fixed computational effort);
3. robustness with respect to physical conditions: the computational cost for a given accuracy should be essentially insensitive to change of physical parameters.

We aim to develop methods fulfilling the above quality criteria for physical models which all display a dissipative behavior, and that are motivated by industrial collaborations or multidisciplinary projects. In particular, we have identified a couple of specific situations we plan to investigate: models from corrosion science (in the framework of nuclear waste repository) [1], low-frequency electromagnetism [10], and mechanics of complex inhomogeneous fluids arising in avalanches [6] or in porous media [46].

Ideally, we should allow ourselves to design entirely new numerical methods. For some applications however (often in the context of industrial collaborations), the members of the team have to compose with existing codes. The numerical algorithms have thus to be optimized under this constraint.

2.2. Scientific Context

Some technological bottlenecks related to points (a)–(c) mentioned above are well identified. In particular, it appears that a good numerical method should handle general meshes, so that dynamic mesh adaptation strategies can be used in order to achieve (b). But it should also be of the highest possible order while remaining stable in the sense of (a), and robust in the sense of (c). There have been numerous research contributions on each point of (a)–(c) in the last decades, in particular for solving each difficulty appart, but combining them still leads to unsolved problems of crucial interest.

Let us mention for example the review paper by J. Droniou [58], where it is highlighted that all the linear methods for solving diffusion equations on general meshes suffer from the same lack of monotonicity and preserve neither the positivity of the solutions nor the decay of the entropy. Moreover, there is no complete convergence proof for the nonlinear methods exposed in [58]. The first convergence proof for a positivity preserving and entropy diminishing method designed to approximate transient dissipative equation on general meshes was proposed very recently in [16]. The idea and the techniques introduced in [16] should be extended to practical applications.

In systems of PDEs, the values of physical parameters often change the qualitative behavior of the solution. Then, one challenge in the numerical approximation of such systems is the design of methods which can be applied for a large range of parameters, as in particular in the regime of singular perturbations. Such schemes, called *asymptotic-preserving* (AP) schemes [65], are powerful tools as they permit the use of the same scheme for a given problem and for its limit with fixed discretization parameters. In many cases, the AP property of numerical schemes is just empirically established, without any rigorous proof. We aim to extend the techniques recently introduced in [3] for the drift-diffusion system, and based on the control of the numerical dissipation of entropy, to other dissipative systems in order prove the AP property of numerical schemes.

The question of the robustness of the numerical methods with respect to the physical parameters is also fundamental for fluid mixtures models. The team already developed such schemes for the variable density Navier-Stokes system [5] or [6]. We aim to propose new ones for more complex models with the same philosophy in mind. On the one hand, we will be interested in high-order schemes, which must be as simple as possible in view of 3D practical implementations. Let us stress that combining high order accuracy and stability is very challenging. On the other hand, the optimization of the computation will have to be considered, in particular with the development of some *a posteriori error* estimators. Impressive progresses have been achieved in this field [54], allowing important computational savings without compromising the accuracy of the results. Recently, we successfully applied this strategy to the Reissner-Mindlin model arising in solid mechanics [53], the dead-oil model for porous media flows [50] or the Maxwell equations in their quasi-static approximation for some eddy current problems [10] and [52]. We aim to develop new *a posteriori* estimators for other dissipative systems, like fluid mixtures models.

In a nutshell, our goal is to take advantage of and extend the most recent breakthroughs of the mathematical community to tackle in an efficient way some application-guided problems coming either from academics or from industrial partners. To this end, we shall focus on the following objectives, which are necessary for the applications we have in mind:

1. *Design and numerical analysis of structure preserving numerical methods*.
2. *Computational optimization*.

3. Research Program

3.1. Design and analysis of structure preserving schemes

3.1.1. Numerical analysis of nonlinear numerical methods

Up to now, the numerical methods dedicated to degenerate parabolic problems that the mathematicians are able to analyze almost all rely on the use of mathematical transformations (like e.g. the Kirchhoff's transform). It forbids the extension of the analysis to complex realistic models. The methods used in the industrial codes for solving such complex problems rely on the use of what we call NNM, i.e., on methods that preserve all the nonlinearities of the problem without reducing them thanks to artificial mathematical transforms. Our aim is to take advantage on the recent breakthrough proposed by C. Cancès & C. Guichard [16], [30] to develop efficient new numerical methods with a full numerical analysis (stability, convergence, error estimates, robustness w.r.t. physical parameters, ...).

3.1.2. Design and analysis of asymptotic preserving schemes

There has been an extensive effort in the recent years to develop numerical methods for diffusion equations that are robust with respect to heterogeneities, anisotropy, and the mesh (see for instance [58] for an extensive discussion on such methods). On the other hand, the understanding of the role of nonlinear stability properties in the asymptotic behaviors of dissipative systems increased significantly in the last decades (see for instance [51], [72]).

Recently, C. Chainais-Hillairet and co-authors [3], [8] and [19] developed a strategy based on the control of the numerical counterpart of the physical entropy to develop and analyze AP numerical methods. In particular, these methods show great promises for capturing accurately the behavior of the solutions to dissipative problems when some physical parameter is small with respect to the discretization characteristic parameters, or in the long-time asymptotic. Since it requires the use of nonlinear test functions in the analysis, strong restrictions on the physics (isotropic problems) and on the mesh (Cartesian grids, Voronoï boxes...) are required in [3], [8] and [19]. The schemes proposed in [16], [30] allow to handle nonlinear test functions in the analysis without restrictions on the mesh and on the anisotropy of the problem. Combining the nonlinear schemes *à la* [16] with the methodology of [3], [8], [19] would provide schemes that are robust both with respect to the meshes and to the parameters. Therefore, they would be also robust under adaptive mesh refinement.

3.1.3. Design and stability analysis of numerical methods for mixture problems

We aim at extending the range of the NS2DDV-M software by introducing new physical models, like for instance the Kazhikov and Smagulov model [68]. This will require a theoretical study for proving the existence of weak solutions to this model. Then, we will need to design numerical schemes to approximate these models and study their stability. We will also study their convergence following the path proposed in [62], [69].

3.2. Optimizing the computational efficiency

3.2.1. High order nonlinear numerical methods

The numerical experiments carried out in [16] show that in case of very strong anisotropy, the convergence of the proposed NNM becomes too slow (less than first order). Indeed, the method appears to strongly overestimate the dissipation. In order to make the method more competitive, it is necessary to estimate the dissipation in a more accurate way. Preliminary numerical results show that second order accuracy in space can be achieved in this way. One also aims to obtain (at least) second order accuracy in time without jeopardizing the stability. For many problems, this can be done by using so-called two-step backward differentiation formulas (BDF2) [59].

Concerning the inhomogeneous fluid models, we aim to investigate new methods for the mass equation resolution. Indeed, we aim at increasing the accuracy while maintaining some positivity-like properties and the efficiency for a wide range of physical parameters. To this end, we will consider *residual distribution* (RD) schemes, that appear as an alternative to finite volume methods. RD schemes enjoy very compact stencils. Therefore, their extension from 2D to 3D yield reasonable difficulties. These methods appeared twenty years ago, but recent extensions to unsteady problems [73], [64], with high-order accuracy [40], [39], or for parabolic problems [37], [38] make them very competitive. Relying on these breakthroughs, we aim at designing new RD schemes for fluid mixture models with high-order accuracy while preserving the positivity of the solutions.

3.2.2. A posteriori error control

The question of the *a posteriori* error estimators will also have to be addressed in this optimization context. Since the pioneering papers of Babuska and Rheinboldt more than thirty years ago [44], *a posteriori* error estimators have been widely studied. We will take advantage of the huge corresponding bibliography database in order to optimize our numerical results.

For example, we would like to generalize the results we derived for the harmonic magnetodynamic case (e.g. [10] and [52]) to the temporal magnetodynamic one, for which space/time *a posteriori* error estimators have to be developed. A space/time refinement algorithm should consequently be proposed and tested on academic as well as industrial benchmarks.

We also want to develop *a posteriori* estimators for the variable density Navier-Stokes model or some of its variants. To do so, several difficulties have to be tackled: the problem is nonlinear, unsteady, and the numerical method [5], [6] we developed combines features from finite elements and finite volumes. Fortunately, we do not start from scratch. Some recent references are devoted to the unsteady Navier-Stokes model in the finite element context [47], [77]. In the finite volume context, recent references deal with unsteady convection-diffusion equations [76], [43], [57] and [50]. We want to adapt some of these results to the variable density Navier-Stokes system, and to be able to design an efficient space-time remeshing algorithm.

3.2.3. Efficient computation of pairwise interactions in large systems of particles

Many systems are modeled as a large number of punctual individuals (N) which interact pairwise which means $N(N-1)/2$ interactions. Such systems are ubiquitous, they are found in chemistry (Van der Waals interaction between atoms), in astrophysics (gravitational interactions between stars, galaxies or galaxy clusters), in biology (flocking behavior of birds, swarming of fishes) or in the description of crowd motions. Building on the special structure of convolution type of the interactions, the team develops computation methods based on the Non Uniform Fast Fourier Transform [61]. This reduces the $O(N^2)$ naïve computational cost of the interactions to $O(N \log N)$, allowing numerical simulations involving millions of individuals.

4. Application Domains

4.1. Porous media flows

Porous media flows are of great interest in many contexts, like, e.g., oil engineering, water resource management, nuclear waste repository management, or carbon dioxyde sequestration. We refer to [46], [45] for an extensive discussion on porous media flow models.

From a mathematical point of view, the transport of complex fluids in porous media often leads to possibly degenerate parabolic conservation laws. The porous rocks can be highly heterogeneous and anisotropic. Moreover, the grids on which one intends to solve numerically the problems are prescribed by the geological data, and might be non-conformal with cells of various shapes. Therefore, the schemes used for simulating such complex flows must be particularly robust.

4.2. Corrosion and concrete carbonatation

The team is interested in the theoretical and numerical analysis of mathematical models describing degradation of materials as concrete carbonation and corrosion. The study of such models is an important environmental and industrial issue. Atmospheric carbonation degrades reinforced concretes and limits the lifetime of civil engineering structures. Corrosion phenomena issues occur for instance in the reliability of nuclear power plants and the nuclear waste repository. The study of the long time evolution of these phenomena is of course fundamental in order to predict the lifetime of the structures.

From a mathematical point of view, the modeling of concrete carbonation (see [41]) as the modeling of corrosion in an underground repository (DPCM model developed by Bataillon *et al.* [1]) lead to systems of PDEs posed on moving domains. The coupling between convection-diffusion-reaction equations and moving boundary equations leads to challenging mathematical questions.

4.3. Complex fluid flows

The team is interested in some numerical methods for the simulation of systems of PDEs describing complex flows, like for instance, mixture flows, granular gases, rarefied gases, or quantum fluids.

Let us first focus on fluid mixture flows. The fluid is described by its density, its velocity and its pressure. These quantities obey mass and momentum conservation. On the one hand, when we deal with the 2D variable density incompressible Navier-Stokes equations, we aim to study the ability of the numerical scheme to reproduce some instabilities phenomena such as the Rayleigh-Taylor instability. On the other hand, diffuse interface models have gained renewed interest for the last few years in fluid mechanics applications. From a physical viewpoint, they allow to describe some phase transition phenomena. If the Fick's law relates the divergence of the velocity field to derivatives of the density, one obtains the so called Kazhikhov-Smagulov model [68]. Here, the density of the mixture is naturally highly non homogeneous, and the constitutive law accounts for diffusion effects between the constituents of the mixture. Models of this type can be used for instance to simulate powder-snow avalanches [6], low-Mach flows, or hydrodynamic models arising in combustion theory or transport of pollutants.

Kinetic theory of molecular gases models a gas as a system of elastically colliding spheres, conserving mechanical energy during impact. Once initialized, it takes a molecular gas not more than few collisions per particle to relax to its equilibrium state, characterized by a Maxwellian velocity distribution and a certain homogeneous density (in the absence of external forces). A granular gas is a system of dissipatively colliding, macroscopic particles (grains). This slight change in the microscopic dynamics (converting energy into heat) cause drastic changes in the behavior of the gas: granular gases are open systems, which exhibits self-organized spatio-temporal cluster formations, and has no equilibrium distribution. They can be used to model silos, avalanches, pollen or planetary rings.

The quantum models can be used to describe superfluids, quantum semiconductors, weakly interacting Bose gases or quantum trajectories of Bohmian mechanics. They have attracted considerable attention in the last decades, due in particular to the development of the nanotechnology applications. To describe quantum phenomena, there exists a large variety of models. In particular there exist three different levels of description: microscopic, mesoscopic and macroscopic. The quantum Navier-Stokes equations deal with a macroscopic description in which the quantum effects are taken into account through a third order term called the quantum Bohm potential. This Bohm potential arises from the fluid dynamical formulation of the single-state Schrödinger equation. The non-locality of quantum mechanics is approximated by the fact that the equations of state do not only depend on the particle density but also on its gradient. These equations were employed to model field emissions from metals and steady-state tunneling in metal- insulator- metal structures and to simulate ultra-small semiconductor devices.

4.4. Stratigraphy

The knowledge of the geology is a prerequisite before simulating flows within the subsoil. Numerical simulations of the geological history thanks to stratigraphy numerical codes allow to complete the knowledge of the geology where experimental data are lacking. Stratigraphic models consist in a description of the erosion and sedimentation phenomena at geological scales.

The characteristic time scales for the sediments are much larger than the characteristic time scales for the water in the river. However, the (time-averaged) water flux plays a crucial role in the evolution of the stratigraphy. Therefore, defining appropriate models that take the coupling between the rivers and the sediments into account is fundamental and challenging. Once the models are at hand, efficient numerical methods must be developed.

4.5. Low frequency electromagnetism

Numerical simulation is nowadays an essential tool in order to design electromagnetic systems, by estimating the electromagnetic fields generated in a wide variety of devices. An important challenge for many applications is to quantify the intensity of the electric field induced in a conductor by a current generated in its neighborhood. In the low-frequency regime, we can for example quote the study of the impact on the human body of a high-tension line or, for higher frequencies, the one of a smartphone. But the ability to simulate accurately some electromagnetic fields is also very useful for non destructive control, in the context of the maintenance of nuclear power stations for example. The development of efficient numerical tools, among which the so-called“*a posteriori* error estimators", is consequently necessary to reach a high precision of calculations in order to provide estimations as reliable as possible.

5. Highlights of the Year

5.1. Highlights of the Year

The paper [31], written by Giacomo Dimarco, Raphaël Loubère, Jacek Narski and Thomas Rey presents a new deterministic numerical scheme for the resolution of the full 7d Boltzmann equation. The scheme combines a robust and fast method for treating the transport part based on an innovative Lagrangian technique supplemented with fast spectral solvers to treat the collision operator. This approach along with several implementation features related to the parallelization of the algorithm permits to construct an efficient simulation tool which is numerically tested against exact and reference solutions on classical problems arising in rarefied gas dynamics.

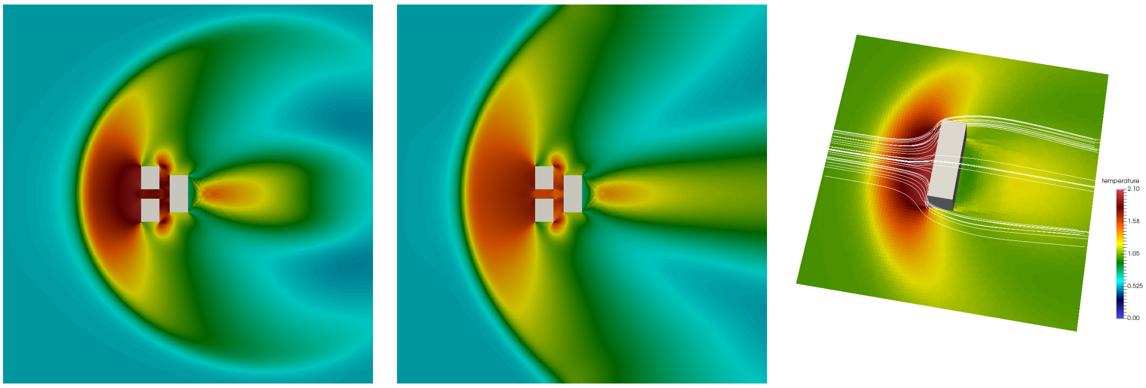


Figure 1. Simulation of a space shuttle atmospheric reentry (pictures from [31])

The paper presents results up to the very challenging 3D×3D case for unsteady flows arising during a space shuttle atmospheric reentry (which was simulated in the deterministic case in the paper for the first time up to our knowledge), which may serve as benchmark for future comparisons between different numerical methods for solving the multidimensional Boltzmann equation. For this reason, the paper also provide for each problem studied details on the implementation, computational cost and memory consumption as well as comparisons with the more standard BGK model or the limit model of compressible Euler equations.

6. New Software and Platforms

6.1. New Software

6.1.1. NS2DDV-M: a code for the simulation of inhomogeneous fluid flow

We develop and freely distribute a new version of the matlab code NS2DDV-M (equipped with a graphic interface and an accurate documentation) to promote new collaborations in the domain, allow some easy comparisons with concurrent codes on the same benchmark cases, and compare alternative numerical solution methods. Contacts: Caterina Calgaro & Emmanuel Creusé.

6.1.2. A scientific computing software for fast simulation of large systems of interacting particles

Benoît Merlet and Thomas Rey have developed a set of numerical codes for the numerical simulation of large systems of interacting particles. For a system of N particles, the number of interactions is a quadratic functions of N, leading to a quadratic cost of a brut force implementation. This fact limits simulations by "naïve" methods to systems with "only" tenth of thousands of particles. In order to treat larger systems (involving millions of particles), the team has implemented a method based on the Non Uniform Fast Fourier Transform which reduces the computation cost of the interactions to $O(N \log N)$. The NUFFT is used to handle the long range smooth interactions. To treat the possibly singular short range interactions (involving only neighboring particles) a quadtree-like method is used. The method is applied to two kind of problems : computations

of the dynamics of interacting particles where a standard ordinary differential equation is used; numerical optimization of the energy of a system of interacting particles thanks to a Nonlinear Conjugate Gradient method.
As an illustration of the efficiency of the code, the team has performed numerical experiments which support the following crystallization conjecture : in 2D, a large number of identical charged particles tend to arrange themselves into a regular triangular lattice.
A user friendly version will be released to the public in 2017.

6.1.3. The Fast Spectral Kinetic Scheme

The Fast Spectral Kinetic Scheme (FSKS), has been jointly developed by researchers from the universities of Ferrara, Toulouse, and Lille, and is the first high-order 7-dimensional deterministic numerical method capable of dealing with the complete physics of rarefied gas dynamics. The FSKM indeed solves the Boltzmann equation in 1 dimension of time, 3 of physical space and 3 of velocity space, and has been used to model accurately aerospace engineering problems such as space shuttle re-entry in the atmosphere or very rarefied gas flow in microscopic devices (Knudsen pump).

7. New Results

7.1. large-time behavior of some numerical schemes

In [19], C. Chainais-Hillairet, A. Jüngel and S. Schuchnigg prove the time decay of fully discrete finite-volume approximations of porous-medium and fast-diffusion equations with Neumann or periodic boundary conditions in the entropy sense. The algebraic or exponential decay rates are computed explicitly. In particular, the numerical scheme dissipates all zeroth-order entropies which are dissipated by the continuous equation. The proofs are based on novel continuous and discrete generalized Beckner inequalities.

In [13], M. Bessemoulin-Chatard and C. Chainais-Hillairet study the large-time behavior of a numerical scheme discretizing drift-diffusion systems for semiconductors. The numerical method is based on a generalization of the classical Scharfetter-Gummel scheme which allows to consider both linear or nonlinear pressure laws.They study the convergence of approximate solutions towards an approximation of the thermal equilibrium state as time tends to infinity, and obtain a decay rate by controlling the discrete relative entropy with the entropy production. This result is proved under assumptions of existence and uniform-in-time L^∞ estimates for numerical solutions, which are then discussed.

The question of uniform-in-time L^∞ estimates for the scheme proposed in [13] has then be tackled by M. Bessemoulin-Chatard, C. Chainais-Hillairet and A. Jüngel. The result is obtained *via* a Moser's iteration technique adapted to the discrete setting. Related to this question, the existence of a positive lower bound for the numerical solution of a convection-diffusion equation has been studied by C. Chainais-Hillairet, B. Merlet and A. Vasseur. They apply a method due to De Giorgi in order to establish a positive lower bound for the numerical solution of a stationary convection-diffusion equation. These results are submitted for publication in the FVCA8 conference (to be held in June 2017).

In [11] B. Merlet *et al.* consider a second-order two-step time discretization of the Cahn-Hilliard equation with an analytic nonlinearity. They study the long time behavior of the discrete solution and show that if the time-step is chosen small enough, the sequence generated by the scheme converges to a steady state as time tends to infinity. Convergence rates are also provided. This parallels the behavior of the solutions of the non-discretized solutions and shows the reliability of the scheme for long time simulations. The method of proof is based on the Lojasiewicz-Simon inequality and on the study of the pseudo-energy associated with the discretization which is shown to be non-increasing.

7.2. Theoretical and numerical analysis of corrosion models

The Diffusion Poisson Coupled Model [1] is a model of iron based alloy in a nuclear waste repository. It describes the growth of an oxide layer in this framework. The system is made of a Poisson equation on the electrostatic potential and convection-diffusion equations on the densities of charge carriers (electrons, ferric cations and oxygen vacancies). The DPCM model also takes into account the growth of the oxide host lattice and its dissolution, leading to moving boundary equations. Numerical experiments done for the simulation of this model with moving boundaries show the convergence in time towards a pseudo-steady-state. C. Chainais-Hillairet and T. O. Gallouët prove in [18] the existence of pseudo-stationary solutions for some simplified versions of the DPCM model. They also propose a new scheme in order to compute directly this pseudo-steady-state. Numerical experiments show the efficiency of this method.

The modeling of concrete carbonation also leads to a system of partial differential equations posed on a moving domain. C. Chainais-Hillairet, B. Merlet and A. Zurek propose and analyze a finite volume scheme for the concrete carbonation model. They prove the convergence of the sequence of approximate solutions towards a weak solution. Numerical experiments show the order 2 in space of the scheme and illustrate the $\sqrt{t}$ law of propagation of the size of the carbonated zone. This result is submitted for publication.

7.3. Modeling and numerics for porous media flows

In [16], C. Cancès and C. Guichard propose a nonlinear Control Volume Finite Elements method with upwinding in order to solve possibly nonlinear and degenerate parabolic equations. This method was designed in order to preserve at the discrete level the positivity and the nonlinear stability of the solutions. In [25], A. Ait Hammou Oulhaj, C. Cancès, and C. Chainais-Hillairet extend the approach of [16] to the more complex case of Richards equation modeling saturated/unsaturated flows in anisotropic porous media. The additional complexity comes from the fact that convective terms and elliptic degeneracy are considered in [25]. The scheme preserves at the discrete level the nonnegativity and the nonlinear stability of the solutions. Its convergence is rigorously proved, and numerical results are provided in order to illustrate the behavior of the scheme.

In [49], C. Cancès, T. O. Gallouët, and L. Monsaingeon show that the equations governing two-phase flows in porous media have a formal gradient flow structure. The goal of the longer contribution [29] is then twofold. First, it extends the variational interpretation of [49] to the case where an arbitrary number of phases are in competition to flow within a porous medium. Second, we provide rigorous foundations to our claim. More precisely, the convergence of a minimizing movement scheme *à la* Jordan, Kinderlehrer, and Otto [66] is shown in [29], providing by the way a new existence result for multiphase flows in porous media. The result relies on advances tools related to optimal transportation [75], [74].

7.4. Complex fluid flows: modeling, analysis and numerics

The analysis of the Kazhikhov-Smagulov model was given by Bresch at al. [48] (see also reference therein). These authors prove the global existence of weak solution without assuming small data and without any assumption on the diffusivity coefficient. Following the physical experiment given by Joseph [67], we introduce a Korteweg stress tensor in the previous model. The theory of Korteweg considers the possibility that motions can be driven by additional stresses associated with gradients of density. In process of slow diffusion on miscible incompressible fluids, for example water and glycerin, dynamical effects which mimic surface tension can arise in thin mixing layers where the gradients of density are large. In the context of the PhD thesis of Meriem Ezzoug (July 2016, University of Monastir, Tunisia), C. Calgaro and co-authors study a multiphase incompressible fluid model, called the Kazhikhov-Smagulov-Korteweg model. They prove in [14] that this model is globally well posed in a 3D bounded domain.

In [21], P.-E. Jabin and T. Rey investigate the behavior of granular gases in the limit of small Knudsen number, that is very frequent collisions. They deal with the physically relevant strongly inelastic case, in one dimension of space and velocity. The study of such limit, also known as hydrodynamic limit is to give a reduced description of the kinetic equation, using a fluid approximation. They are able to prove the convergence of the particle distribution function toward a monokinetic distribution, whose moments verify the pressureless Euler system. The proof relies on dispersive relations at the kinetic level, which leads to the so-called Oleinik property at the limit, and in particular stability of the solution to the fluid problem.

In [34], I. Lacroix-Violet and A. Vasseur present the construction of global weak solutions to the quantum Navier-Stokes equation, for any initial value with bounded energy and entropy. The construction is uniform with respect to the Planck constant. This allows to perform the semi-classical limit to the associated compressible Navier-Stokes equation. One of the difficulty of the problem is to deal with the degenerate viscosity, together with the lack of integrability on the velocity. The method is based on the construction of weak solutions that are renormalized in the velocity variable. The existence, and stability of these solutions do not need the Mellet-Vasseur inequality [71].

In [31], G. Dimarco, R. Loubère, J. Narski and T. Rey deal with the extension of the Fast Kinetic Scheme (FKS) [55], [56] originally constructed for solving the BGK equation, to the more challenging case of the Boltzmann equation. The scheme combines a robust and fast method for treating the transport part based on an innovative Lagrangian technique supplemented with fast spectral schemes to treat the collisional operator by means of an operator splitting approach. This approach along with several implementation features related to the parallelization of the algorithm permits to construct an efficient simulation tool which is numerically tested against exact and reference solutions on classical problems arising in rarefied gas dynamic. They present results up to the 3D×3D case for unsteady flows for the Variable Hard Sphere model which may serve as benchmark. For this reason, they also provide for each problem details on the computational cost and memory consumption as well as comparisons with the BGK model or the limit model of compressible Euler equations.

7.5. Improving the numerical efficiency of numerical methods

In this section, we gather contributions in which a methodology was introduced in order to reduce the computational cost at fixed accuracy or to improve the accuracy for a fixed computational cost.

In [20], E. Creusé and his collaborators generalized some of their previous results on residual a posteriori error estimators for low electromagnetism [10] , [52] to the case where some voltage or current excitation is specified in the model (see e.g. such models in [63], [42]). It consequently led to consider different formulations and to overcome some specific difficulties in order to derive the reliability of the involved estimators.

It is now well accepted that well-balanced schemes are of great interest in order to compute accurate solutions to systems of PDEs (see for instance [60]). In [36], L. Pareschi and T. Rey propose a systematic way to tune classical numerical schemes in order to make them well-balanced and asymptotic preserving. Inspired by micro-macro decomposition methods for kinetic equations, they present a class of schemes which are capable to preserve the steady state solution and achieve high order accuracy for a class of time dependent partial differential equations including nonlinear diffusion equations and kinetic equations. Extension to systems of conservation laws with source terms are also discussed, as well as Total Variation Diminishing preserving properties.

The contribution [26] by K. Brenner and C. Cancès is devoted to the improvement of the behavior of Newton's method when solving degenerate parabolic equations. Such equations are very common for instance in the context of complex porous media flows. In [26], the presentation focuses on Richards equation modeling saturated/unsaturated flows in porous media. The basic idea is the following: Newton's method is not invariant by nonlinear change of variables. The choice of the primary variable then impacts the effective resolution of the nonlinear system provided by the scheme. The idea developed in [26] is then to construct an abstract primary variable to facilitate Newton's method's convergence. This leads to an impressive reduction of the computational cost, a better accuracy in the results and an strong robustness of the method w.r.t. the nonlinearities appearing in the continuous model.

7.6. Variational modeling and analysis

Bose-Einstein condensates are a unique way to observe quantum effects at a (relatively) large scale. The fundamental states of such condensates are obtained as minimizers of a Gross-Pitaievskii functional. In [33],

M. Goldman and B. Merlet consider the case of a two component Bose-Einstein condensate in the strong segregation regime (the energy favors spatial segregation of the two different Boson species). They identify two different regimes in the strong segregation and small healing length limit. In one of these regimes, the relevant limit is an interesting weighted isoperimetric problem which explains some of the numerical simulations of [70].

In [32], B. Merlet *et al.* consider the branched transportation problem in 2D associated with a cost per unit length of the form $1 + \alpha m$ where m denotes the amount of transported mass and $\alpha > 0$ is a fixed parameter (the limit case $\alpha = 0$ corresponds to the classical Steiner problem). Motivated by the numerical approximation of this problem, they introduce a family of functionals ($\{F_\varepsilon\}_{\varepsilon>0}$) which approximate the above branched transport energy. They justify rigorously the approximation by establishing the equicoercivity and the Γ-convergence of F_ε as $\varepsilon \downarrow 0$. These functionals are modeled on the Ambrosio-Tortorelli functional and are easy to optimize in practice (the algorithm amounts to perform repetitively the alternate optimization of two quadratic functionals). Numerical evidences of the efficiency of the method are presented.

7.7. Miscellaneous

This section gathers results from members of the team that are not directly related to the core of the scientific program of the team.

In [12], I. Violet-Lacroix and co-authors consider the derivation of continuous and fully discrete artificial boundary conditions for the linearized Korteweg-de-Vries equation. They are provided for two different numerical schemes. The boundary conditions being nonlocal with respect to time variable, they propose fast evaluations of discrete convolutions. Various numerical tests are presented to show the effectiveness of the constructed artificial boundary conditions.

A semi-discrete in time Crank-Nicolson scheme to discretize a weakly damped forced nonlinear fractional Schrödinger equation in the whole space ($\mathbb{R}$ is considered by C. Calgaro and co-authors in [28]. They prove that such semi-discrete equation provides a discrete infinite dimensional dynamical in $H^\alpha(\mathbb{R})$ that possesses a global attractor. They show also that if the external force is in a suitable weighted Lebesgue space then this global attractor has a finite fractal dimension.

In [35], F. Nabet considers a finite-volume approximation, based on a two point flux approximation, for the Cahn-Hilliard equation with dynamic boundary conditions. An error estimate for the fully-discrete scheme on a possibly smooth non-polygonal domain is proved and numerical simulations which validate the theoretical result are given.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

C. Cancès supervises the PhD Thesis of Nicolas Peton at IFPEN since October 15, 2015. The bilateral contract enters the framework agreement between Inria and IFPEN.

9. Partnerships and Cooperations

9.1. Regional Initiatives

The PhD program of Ahmed Aït Hammou Oulhaj is partially supported (50%) by the Region Nord-Pas-de-Calais.

9.2. National Initiatives

9.2.1. ANR

C. Cancès is the coordinator of the ANR GEOPOR project. (https://www.ljll.math.upmc.fr/cances/ANR-GEOPOR/). This project aims to study realistic models for complex porous media flows from a variational point of view, and to take advantage of this new approach to design and analyze some efficient numerical methods.

Title: Approche géométrique pour les écoulements en milieux poreux : théorie et numérique.

Type: Jeunes Chercheuses Jeunes Chercheurs SIMI 1- 2013

ANR Reference: ANR-13-JS01-0007-01

Coordinator: Clément Cancès, Inria Lille - Nord Europe.

Duration: January 2014 – June 2017

I. Lacroix is the local coordinator at Université Lille 1 of the ANR BECASIM project (http://becasim.math.cnrs.fr/). This ANR project gathers mathematicians with theoretical and numerical backgrounds together with engineers. The objective is to develop numerical methods to accurately simulate the behavior of Bose-Einstein condensates.

Title: Simulation numérique avancée pour les condensats de Bose-Einstein.

Type: Modèles Numériques - 2012

ANR reference: ANR-12-MONU-0007

Coordinator: Ionut DANAILA, Université de Rouen.

Duration: January 2013 - November 2017.

C. Chainais-Hillairet is a member of the ANR MOONRISE project (http://moonrise.math.cnrs.fr/). The MOONRISE project aims at exploring modeling, mathematical and numerical issues originating from the presence of high oscillations in nonlinear PDEs mainly from the physics of nanotechnologies and from the physics of plasmas.

Title: Modèles, Oscillations et schémas numériques.

Type: Fondements du numérique (DS0705) - 2014

ANR reference: ANR-14-CE23-0007

Coordinator: Florian MEHATS, Université de Rennes 1.

Duration: October 2014 - September 2019.

B. Merlet is a member of the ANR GEOMETRYA project (https://www.ljll.math.upmc.fr/lemenant/GEOMETRYA/) The GEOMETRYA project addresses several problems within the framework of geometric measure theory, from both theoretical and numerical viewpoints. Most of these problems are derived from the modeling of physical phenomenons. The main topics are: the Geometric Measure Theory in singular metric spaces, the Plateau problem, the Mumford-Shah functional, irrigation and branched transport problems, the Willmore energy.

Title: Théorie gémométrique de la mesure et applications

Type: Blanc SIMI 1 - 2012

ANR reference: ANR-12-BS01-0014

Coordinator: Hervé Pajot, Université Joseph Fourier (Grenoble).

Duration: january 2013 - december 2016.

9.2.2. Labex CEMPI

Title: Centre Européen pour les Mathématiques, la Physique et leurs interactions

Coordinator: Stephan De Bièvre.

Duration: January 2012 - December 2019.

Partners: Laboratoire Paul Painlevé and Laser physics department (PhLAM), Université Lille 1.

The “Laboratoire d’Excellence” Centre Européen pour les Mathématiques, la Physique et leurs interactions (CEMPI), a project of the Laboratoire de Mathématiques Paul Painlevé and the Laboratoire de Physique des Lasers, Atomes et Molécules (PhLAM), was created in the context of the "Programme d’Investissements d’Avenir" in February 2012.

The association Painlevé-PhLAM creates in Lille a research unit for fundamental and applied research and for training and technological development that covers a wide spectrum of knowledge stretching from pure and applied mathematics to experimental and applied physics.

One of the three focus areas of CEMPI research is the interface between mathematics and physics. This focus area encompasses three themes. The first is concerned with key problems of a mathematical, physical and technological nature coming from the study of complex behavior in cold atoms physics and non-linear optics, in particular fibre optics. The two other themes deal with fields of mathematics such as algebraic geometry, modular forms, operator algebras, harmonic analysis and quantum groups that have promising interactions with several branches of theoretical physics.

9.3. International Research Visitors

9.3.1. Visits of International Scientists

Alexis Vasseur (UT Austin, Texas) was invited in Lille in June 2016 thanks to a support of the Labex CEMPI.

We have a long time collaboration with Ansgar Jüngel’s team from TU Wien. In 2016, we hosted 2 PhD students advised by A. Jüngel : Anita Gerstenmayer for a first one month and a second one week research stays, Polina Shpartko for a one week research stay.

Kyle Talbot, a PhD student advised by Jérôme Droniou at Monash University (Melbourne, Australia), and Ward Melis, a PhD student supervised by Giovanni Samaey (KU Leuven, Belgium), spent both one week in our team.

9.3.2. Visits to International Teams

Claire Chainais-Hillairet and Ingrid Lacroix-Violet visited Ansgar Jüngel in Vienna (May 17-20, 2016). Claire Chainais-Hillairet visited Jürgen Fuhrmann, Patricio Farrell and Nella Rotundo at WIAS (Berlin) to work on numerical schemes for semiconductor devices models. Clément Cancès visited Léonard Monsaingeon in Lisbon (Feb. 29 to March 4, 2016) Clément Cancès and Flore Nabet visited Daniel Matthes in Munich (June 6-8).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

The team organized the second edition of the conference ABPDE *Asymptotic Behavior of systems of PDE arising in physics and biology: theoretical and numerical points of view* in Lille (June 15-17, 2016). See https://indico.math.cnrs.fr/event/939/.

The kick-off meeting of the GdR MaNu was organized by Clément Cancès, Corinne Jamroz, and Nicolas Seguin (Univ. Rennes) in Saint-Valery-sur-Somme in October 2016. See https://indico.math.cnrs.fr/event/1575/.

Claire Chainais-Hillairet and Clément Cancès are members of the organizing committee of the eighth symposium on Finite Volumes for Complex Applications (FVCA8) to be held in Lille next June 2017. See https://indico.math.cnrs.fr/event/1299/.

I. Lacroix-Violet is in charge of the organization of the weekly seminary of the Numerical Analysis and Partial Differential Equations (ANEDP) research team at the Laboratoire Paul Painlevé, Université de Lille 1.

10.1.2. Journal

10.1.2.1. Member of the editorial boards

C. Chainais-Hillairet is a member of the editorial board of the North-Western European Journal of Mathematics (`http://math.univ-lille1.fr/ nwejm/`) and of the International Journal on Finite Volumes (`http://www.i2m.univ-amu.fr/IJFV/`).

10.1.2.2. Reviewer - Reviewing activities

The members of the team RAPSODI reviewed numerous papers for numerous international journals.

10.1.3. Invited Talks

C. Cancès was an invited speaker at the international conference *Advanced numerical methods: recent developments, analysis, and applications* held in the framework of the IHP quarter on *Numerical Methods for PDEs*.

10.1.4. Scientific Expertise

C. Chainais-Hillairet and E. Creusé were experts for the HCERES.

10.1.5. Research Administration

Clément Cancès is the head of the MaNu Research Group (GdR MaNu, http://gdr-manu.math.cnrs.fr/) funded by the Institute for Mathematical Sciences and Interaction (INSMI) of the French National Center for Research (CNRS).

E. Creusé is AMIES Facilitator (Agency for the Interaction of Mathematics with Enterprise and Society) for the Northern France area. He is also the industrial representative of the Paul Painlevé Laboratory.

C. Chainais-Hillairet is head of the Commission Emplois de Recherche of the Lille - Nord Europe Inria research center.

Caterina Calgaro is a member of the Commission de la Formation et de la Vie Universitaire of the Academic Council of Université Lille 1.

Ingrid Lacroix-Violet, Benoît Merlet and Thomas Rey are members of the Conseil du Laboratoire Paul Painlevé.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

The group is strongly involved in teaching at the Université Lille 1. C. Calgaro and C. Chainais-Hillairet are in charge respectively of the Master of Mathematical Engineering and of the Master 2 of Scientific Computing, whereas E. Creusé is responsible of the "Cursus Master Ingénierie" in Mathematics, Lille 1 University. C. Cancès gives lectures at Polytech' UPMC.

10.2.2. Supervision

PhD : Pierre-Louis Colin has defended his PhD thesis on June 27, 2016. *Analyse numérique de modèles de dérive-diffusion : convergence et comportements asymptotiques*, Univ. Lille 1. advisors: C. Chainais-Hillairet and I. Lacroix-Violet.

PhD : Roberta Tittarelli, *A posteriori error estimators for Maxwell equations in potential and temporal formulations*, Univ. Lille 1, defended on September 27, 2016, advisors: E. Creusé & F. Piriou.

PhD in progress: Ahmed Aït Hammou Oulhaj, *Design and analysis of nonlinear numerical schemes for solving parabolic problems: application to porous media flows*, since 01/10/2014, advisors: C. Cancès & C. Chainais-Hillairet.

PhD in progress: Claire Colin, *Analyse numérique et simulations de modèles multifluides*, since 01/10/2015, advisors: C. Calgaro & E. Creusé.

PhD in progress: Luca Ferrari, *Line energies and applications to image reconstruc- tion of partially masked objects*, since 01/09/2015, advisors: A. Chambolle (CNRS & CMAP, École Polytechnique) & B. Merlet.

PhD in progress: Nicolas Peton, *Numerical methods for a stratigraphic model with nonlinear diffusion and moving frontier areas*, 15/10/2015, C. Cancès, Q. H. Tran (IFPEN) & S. Wolf (IFPEN).

PhD in progress: Antoine Zurek, *Numerical and theoretical analysis of models describing the corrosion of materials*, since 01/10/2016, advisors: C. Chainais-Hillairet & B. Merlet.

10.2.3. Juries

C. Cancès reported on Mayya Groza's PhD thesis, defended on November 10, 2016 at Univ. Nice - Sophia Antipolis. Title: *Modélisation et discrétisation des écoulements diphasiques en milieux poreux avec réseaux de fractures discrètes*

C. Chainais-Hillairet reported on Polina Shpartko's PhD thesis, defended on June 6, 2016 at TU Wien. Title: *Analytical and numerical study of drift-diffusion models for spin-transport in semiconductors*. She was also a member of the jury of Toko Kamtchueng's PhD thesis, defended on December 7, 2016 at Université d'Orléans. Title: *Formulation généralisée du transport réactif pour les modèles de réseaux de pores saturés en eau*.

E. Creusé reported on Azba Riaz' PhD thesis, defended on April 4, 2016 at Univ. Cergy-Pontoise. Title: *A new discontinuous Galerkin formulation for time dependant Maxwell's equation : a priori and a posteriori error estimation*. He was also a member of the jury of Florent Dewez' PhD thesis defended on November 3, 2016 at Univ. Lille 1. Title: *Estimations sans pertes pour des méthodes asymptotiques et notion de propagation pour des équations dispersives.*

B. Merlet reported François Dayrens' PhD thesis, defended in july 1st, 2016 at University Lyon 1. Title: *Minimizing movement and gradient flows for second order geometric functionals*.

10.3. Popularization

C. Calgaro is in charge of the communication of "Laboratoire Paul Painlevé" and she is in charge of the relation between the University of Lille 1 and high schools. Accordingly, she organizes various events which promote mathematics among young peoples like

Les Mathématiques itinérantes (http://mathematiques.univ-lille1.fr/Ouvertures/Mathematiques-itinerantes/)

La semaine des Mathématiques (http://mathematiques.univ-lille1.fr/Ouvertures/Mathematiques-itinerantes/)

Stage en sciences pour les élèves de seconde (http://www.univ-lille1.fr/etudes/stageseconde)

Members of the team participate regularly in theses actions.

Thomas Rey animates a Mathematics workshop (Math en Jeans) in the Adolphe Delégorgue middle school at Courcelles-lès-Lens.

11. Bibliography

Major publications by the team in recent years

[1] C. BATAILLON, F. BOUCHON, C. CHAINAIS-HILLAIRET, C. DESGRANGES, E. HOARAU, F. MARTIN, S. PERRIN, M. TUPIN, J. TALANDIER. *Corrosion modelling of iron based alloy in nuclear waste repository*, in "Electrochim. Acta", 2010, vol. 55, n^o 15, p. 4451–4467.

[2] C. BATAILLON, F. BOUCHON, C. CHAINAIS-HILLAIRET, J. FUHRMANN, E. HOARAU, R. TOUZANI. *Numerical methods for the simulation of a corrosion model with moving oxide layer*, in "J. Comput. Phys.", 2012, vol. 231, n^o 18, p. 6213–6231, http://dx.doi.org/10.1016/j.jcp.2012.06.005.

[3] M. BESSEMOULIN-CHATARD, C. CHAINAIS-HILLAIRET, M.-H. VIGNAL. *Study of a fully implicit scheme for the drift-diffusion system. Asymptotic behavior in the quasi-neutral limit.*, in "SIAM, J. Numer. Anal.", 2014, vol. 52, n^o 4, http://epubs.siam.org/toc/sjnaam/52/4.

[4] C. CALGARO, E. CHANE-KANE, E. CREUSÉ, T. GOUDON. *L^∞-stability of vertex-based MUSCL finite volume schemes on unstructured grids: simulation of incompressible flows with high density ratios*, in "J. Comput. Phys.", 2010, vol. 229, n^o 17, p. 6027–6046.

[5] C. CALGARO, E. CREUSÉ, T. GOUDON. *An hybrid finite volume-finite element method for variable density incompressible flows*, in "J. Comput. Phys.", 2008, vol. 227, n^o 9, p. 4671–4696.

[6] C. CALGARO, E. CREUSÉ, T. GOUDON. *Modeling and simulation of mixture flows: application to powder-snow avalanches*, in "Comput. & Fluids", 2015, vol. 107, p. 100–122, http://dx.doi.org/10.1016/j.compfluid.2014.10.008.

[7] C. CANCÈS, C. GUICHARD. *Convergence of a nonlinear entropy diminishing Control Volume Finite Element scheme for solving anisotropic degenerate parabolic equations*, in "Mathematics of Computation", 2016, vol. 85, n^o 298, p. 549-580, https://hal.archives-ouvertes.fr/hal-00955091.

[8] C. CHAINAIS-HILLAIRET. *Entropy method and asymptotic behaviours of finite volume schemes*, in "Finite volumes for complex applications. VII. Methods and theoretical aspects", Springer Proc. Math. Stat., Springer, Cham, 2014, vol. 77, p. 17–35.

[9] E. CREUSÉ, S. NICAISE, G. KUNERT. *A posteriori error estimation for the Stokes problem: anisotropic and isotropic discretizations*, in "Math. Models Methods Appl. Sci.", 2004, vol. 14, n^o 9, p. 1297–1341, http://dx.doi.org/10.1142/S0218202504003635.

[10] E. CREUSÉ, S. NICAISE, Z. TANG, Y. LE MENACH, N. NEMITZ, F. PIRIOU. *Residual-based a posteriori estimators for the $\mathbf{A}-\phi$ magnetodynamic harmonic formulation of the Maxwell system*, in "Math. Models Methods Appl. Sci.", 2012, vol. 22, n^o 5, 1150028, 30, http://dx.doi.org/10.1142/S021820251150028X.

Publications of the year

Articles in International Peer-Reviewed Journal

[11] P. F. ANTONIETTI, B. MERLET, M. PIERRE, M. VERANI.*Convergence to equilibrium for a second-order time semi-discretization of the Cahn-Hilliard equation*, in "AIMS Mathematics", August 2016, vol. 1, n^{o} 3, p. 178-194, https://hal.archives-ouvertes.fr/hal-01355956.

[12] C. BESSE, M. EHRHARDT, I. LACROIX-VIOLET.*Discrete Artificial Boundary Conditions for the Korteweg-de Vries Equation*, in "Numerical Methods for Partial Differential Equations", 2016, vol. 35, n^{o} 5, p. 1455-1484, https://hal.archives-ouvertes.fr/hal-01105043.

[13] M. BESSEMOULIN-CHATARD, C. CHAINAIS-HILLAIRET.*Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems*, in "Journal of Numerical Mathematics", July 2016, https://hal.archives-ouvertes.fr/hal-01250709.

[14] C. CALGARO, M. EZZOUG, E. ZAHROUNI.*On the global existence of weak solution for a multiphasic incompressible fluid model with Korteweg stress*, in "Mathematical Methods in the Applied Sciences", 2016 [*DOI :* 10.1002/MMA.3969], https://hal.archives-ouvertes.fr/hal-01388718.

[15] C. CANCÈS, F. COQUEL, E. GODLEWSKI, H. MATHIS, N. SEGUIN.*Error analysis of a dynamic model adaptation procedure for nonlinear hyperbolic equations*, in "Communications in Mathematical Sciences", 2016, vol. 14, n^{o} 1, p. 1-30, https://hal.archives-ouvertes.fr/hal-00852101.

[16] C. CANCÈS, C. GUICHARD.*Convergence of a nonlinear entropy diminishing Control Volume Finite Element scheme for solving anisotropic degenerate parabolic equations*, in "Mathematics of Computation", 2016, vol. 85, n^{o} 298, p. 549-580, https://hal.archives-ouvertes.fr/hal-00955091.

[17] C. CANCÈS, H. MATHIS, N. SEGUIN.*Error estimate for time-explicit finite volume approximation of strong solutions to systems of conservation laws*, in "SIAM Journal on Numerical Analysis", 2016, vol. 54, n^{o} 2, p. 1263-1287, https://hal.archives-ouvertes.fr/hal-00798287.

[18] C. CHAINAIS-HILLAIRET, T. GALLOUËT.*Study of a pseudo-stationary state for a corrosion model: existence and numerical approximation*, in "Nonlinear Analysis: Real World Applications", 2016, https://hal.inria.fr/hal-01147621.

[19] C. CHAINAIS-HILLAIRET, A. JÜNGEL, S. SCHUCHNIGG.*Entropy-dissipative discretization of nonlinear diffusion equations and discrete Beckner inequalities*, in "Modelisation Mathématique et Analyse Numérique", 2016, vol. 50, n^{o} 1, p. 135-162, https://hal.archives-ouvertes.fr/hal-00924282.

[20] C. CHEN, E. CREUSÉ, S. NICAISE, Z. TANG.*Residual-based a posteriori estimators for the potential formulations of electrostatic and time-harmonic eddy current problems with voltage or current excitation*, in "International Journal for Numerical Methods in Engineering", 2016, vol. 107, n^{o} 5, 18, https://hal.inria.fr/hal-01390241.

[21] P.-E. JABIN, T. REY.*Hydrodynamic limit of granular gases to pressureless Euler in dimension 1*, in "Quarterly of Applied Mathematics", July 2016, 26, 23 pages, 1 figure [*DOI :* 10.1090/QAM/1442.], https://hal.archives-ouvertes.fr/hal-01279961.

[22] T. Rey, C. Tan.*An Exact Rescaling Velocity Method for some Kinetic Flocking Models*, in "SIAM Journal on Numerical Analysis", 2016, vol. 54, n^{o} 2, p. 641–664, 21 pages, 6 figures [*DOI :* 10.1137/140993430.], https://hal.archives-ouvertes.fr/hal-01078298.

[23] R. Tittarelli, Y. Le Menach, E. Creusé, S. Nicaise, F. Piriou, O. Moreau, O. Boiteau.*Space-time residual-based a posteriori estimator for the A-phi formulation in eddy current problems*, in "IEEE Transactions on Magnetics", 2016, vol. 51, n^{o} 3, https://hal.inria.fr/hal-01390250.

Other Publications

[24] A. Ait Hammou Oulhaj.*Numerical analysis of a finite volume scheme for a seawater intrusion model with cross-diffusion in an unconfined aquifer*, 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01432197.

[25] A. Ait Hammou Oulhaj, C. Cancès, C. Chainais-Hillairet.*Numerical analysis of a nonlinearly stable and positive Control Volume Finite Element scheme for Richards equation with anisotropy*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01372954.

[26] K. Brenner, C. Cancès.*Improving Newton's method performance by parametrization: the case of Richards equation*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01342386.

[27] C. Calgaro, E. Creusé, T. Goudon, S. Krell.*Simulations of non homogeneous viscous flows with incompressibility constraints*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01246070.

[28] C. Calgaro, O. Goubet, E. Zahrouni.*Finite dimensional global attractor for a semi-discrete fractional nonlinear Schrödinger equation*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01388788.

[29] C. Cancès, T. Gallouët, L. Monsaingeon.*Incompressible immiscible multiphase flows in porous media: a variational approach*, 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01345438.

[30] C. Cancès, C. Guichard.*Numerical analysis of a robust free energy diminishing Finite Volume scheme for parabolic equations with gradient structure*, 2016, to appear in Foundations of Computational Mathematics, https://hal.archives-ouvertes.fr/hal-01119735.

[31] G. Dimarco, R. Loubère, J. Narski, T. Rey.*An efficient numerical method for solving the Boltzmann equation in multidimensions*, August 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01357112.

[32] L. A. D. Ferrari, A. Chambolle, B. Merlet.*A simple phase-field approximation of the Steiner problem in dimension two*, September 2016, 24 pages, 8 figures, https://hal.archives-ouvertes.fr/hal-01359483.

[33] M. Goldman, B. Merlet.*Phase segregation for binary mixtures of Bose-Einstein Condensates*, November 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01155676.

[34] I. Lacroix-Violet, A. Vasseur.*Global weak solutions to the compressible quantum navier-stokes equation and its semi-classical limit*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01347943.

[35] F. NABET.*An error estimate for a finite-volume scheme for the Cahn-Hilliard equation with dynamic boundary conditions*, February 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01273945.

[36] L. PARESCHI, T. REY.*Residual equilibrium schemes for time dependent partial differential equations*, February 2016, 23 pages, 12 figures, https://hal.archives-ouvertes.fr/hal-01270297.

References in notes

[37] R. ABGRALL.*A review of residual distribution schemes for hyperbolic and parabolic problems: the July 2010 state of the art*, in "Commun. Comput. Phys.", 2012, vol. 11, n^{o} 4, p. 1043–1080, http://dx.doi.org/10.4208/cicp.270710.130711s.

[38] R. ABGRALL, G. BAURIN, A. KRUST, D. DE SANTIS, M. RICCHIUTO.*Numerical approximation of parabolic problems by residual distribution schemes*, in "Internat. J. Numer. Methods Fluids", 2013, vol. 71, n^{o} 9, p. 1191–1206, http://dx.doi.org/10.1002/fld.3710.

[39] R. ABGRALL, A. LARAT, M. RICCHIUTO.*Construction of very high order residual distribution schemes for steady inviscid flow problems on hybrid unstructured meshes*, in "J. Comput. Phys.", 2011, vol. 230, n^{o} 11, p. 4103–4136, http://dx.doi.org/10.1016/j.jcp.2010.07.035.

[40] R. ABGRALL, A. LARAT, M. RICCHIUTO, C. TAVÉ.*A simple construction of very high order non-oscillatory compact schemes on unstructured meshes*, in "Comput. & Fluids", 2009, vol. 38, n^{o} 7, p. 1314–1323, http://dx.doi.org/10.1016/j.compfluid.2008.01.031.

[41] T. AIKI, A. MUNTEAN.*A free-boundary problem for concrete carbonation: front nucleation and rigorous justification of the $\sqrt{t}$-law of propagation*, in "Interfaces Free Bound.", 2013, vol. 15, n^{o} 2, p. 167–180, http://dx.doi.org/10.4171/IFB/299.

[42] A. ALONSO RODRÍGUEZ, A. VALLI.*Voltage and current excitation for time-harmonic eddy-current problems*, in "SIAM J. Appl. Math.", 2008, vol. 68, n^{o} 5, p. 1477–1494, http://dx.doi.org/10.1137/070697677.

[43] B. AMAZIANE, A. BERGAM, M. EL OSSMANI, Z. MGHAZLI.*A posteriori estimators for vertex centred finite volume discretization of a convection-diffusion-reaction equation arising in flow in porous media*, in "Internat. J. Numer. Methods Fluids", 2009, vol. 59, n^{o} 3, p. 259–284, http://dx.doi.org/10.1002/fld.1456.

[44] I. BABUŠKA, W. C. RHEINBOLDT.*Error estimates for adaptive finite element computations*, in "SIAM J. Numer. Anal.", 1978, vol. 15, n^{o} 4, p. 736–754.

[45] J. BEAR, Y. BACHMAT.*Introduction to modeling of transport phenomena in porous media*, Springer, 1990, vol. 4.

[46] J. BEAR.*Dynamic of Fluids in Porous Media*, American Elsevier, New York, 1972.

[47] S. BERRONE, V. GARBERO, M. MARRO.*Numerical simulation of low-Reynolds number flows past rectangular cylinders based on adaptive finite element and finite volume methods*, in "Comput. & Fluids", 2011, vol. 40, p. 92–112, http://dx.doi.org/10.1016/j.compfluid.2010.08.014.

[48] D. BRESCH, E. H. ESSOUFI, M. SY.*Effect of density dependent viscosities on multiphasic incompressible fluid models*, in "J. Math. Fluid Mech.", 2007, vol. 9, n^{o} 3, p. 377–397.

[49] C. CANCÈS, T. O. GALLOUËT, L. MONSAINGEON.*The gradient flow structure for incompressible immiscible two-phase flows in porous media*, in "C. R. Math. Acad. Sci. Paris", 2015, vol. 353, n^{o} 11, p. 985–989, http://dx.doi.org/10.1016/j.crma.2015.09.021.

[50] C. CANCÈS, I. S. POP, M. VOHRALÍK.*An a posteriori error estimate for vertex-centered finite volume discretizations of immiscible incompressible two-phase flow*, in "Math. Comp.", 2014, vol. 83, n^{o} 285, p. 153–188, http://dx.doi.org/10.1090/S0025-5718-2013-02723-8.

[51] J. A. CARRILLO, A. JÜNGEL, P. A. MARKOWICH, G. TOSCANI, A. UNTERREITER.*Entropy dissipation methods for degenerate parabolic problems and generalized Sobolev inequalities*, in "Monatsh. Math.", 2001, vol. 133, n^{o} 1, p. 1–82, http://dx.doi.org/10.1007/s006050170032.

[52] E. CREUSÉ, S. NICAISE, Z. TANG, Y. LE MENACH, N. NEMITZ, F. PIRIOU.*Residual-based a posteriori estimators for the $\mathbf{T}/\Omega$ magnetodynamic harmonic formulation of the Maxwell system*, in "Int. J. Numer. Anal. Model.", 2013, vol. 10, n^{o} 2, p. 411–429.

[53] E. CREUSÉ, S. NICAISE, E. VERHILLE.*Robust equilibrated a posteriori error estimators for the Reissner-Mindlin system*, in "Calcolo", 2011, vol. 48, n^{o} 4, p. 307–335, http://dx.doi.org/10.1007/s10092-011-0042-0.

[54] D. A. DI PIETRO, M. VOHRALÍK.*A Review of Recent Advances in Discretization Methods, a Posteriori Error Analysis, and Adaptive Algorithms for Numerical Modeling in Geosciences*, in "Oil & Gas Science and Technology-Rev. IFP", June 2014, p. 1-29, (online first).

[55] G. DIMARCO, R. LOUBERE.*Towards an ultra efficient kinetic scheme. Part I: Basics on the BGK equation*, in "J. Comput. Phys.", 2013, vol. 255, p. 680–698, http://dx.doi.org/10.1016/j.jcp.2012.10.058.

[56] G. DIMARCO, R. LOUBERE.*Towards an ultra efficient kinetic scheme. Part II: The high order case*, in "J. Comput. Phys.", 2013, vol. 255, p. 699–719, http://dx.doi.org/10.1016/j.jcp.2013.07.017.

[57] V. DOLEJŠÍ, A. ERN, M. VOHRALÍK.*A framework for robust a posteriori error control in unsteady nonlinear advection-diffusion problems*, in "SIAM J. Numer. Anal.", 2013, vol. 51, n^{o} 2, p. 773–793, http://dx.doi.org/10.1137/110859282.

[58] J. DRONIOU.*Finite volume schemes for diffusion equations: introduction to and review of modern methods*, in "Math. Models Methods Appl. Sci.", 2014, vol. 24, n^{o} 8, p. 1575-1620.

[59] E. EMMRICH.*Two-step BDF time discretisation of nonlinear evolution problems governed by monotone operators with strongly continuous perturbations*, in "Comput. Methods Appl. Math.", 2009, vol. 9, n^{o} 1, p. 37–62.

[60] L. GOSSE.*Computing qualitatively correct approximations of balance laws*, SIMAI Springer Series, Springer, Milan, 2013, vol. 2, xx+340, Exponential-fit, well-balanced and asymptotic-preserving, http://dx.doi.org/10.1007/978-88-470-2892-0.

[61] L. GREENGARD, J.-Y. LEE.*Accelerating the nonuniform fast Fourier transform*, in "SIAM Rev.", 2004, vol. 46, n^{o} 3, p. 443–454, http://dx.doi.org/10.1137/S003614450343200X.

[62] F. GUILLÉN-GONZÁLEZ, J. V. GUTIÉRREZ-SANTACREU.*Conditional stability and convergence of a fully discrete scheme for three-dimensional Navier-Stokes equations with mass diffusion*, in "SIAM J. Numer. Anal.", 2008, vol. 46, n^{o} 5, p. 2276–2308, http://dx.doi.org/10.1137/07067951X.

[63] R. HIPTMAIR, O. STERZ.*Current and voltage excitations for the eddy current model*, in "International Journal of Numerical Modelling: Electronic Networks, Devices and Fields", 2005, vol. 18, n^{o} 1, p. 1–21.

[64] M. E. HUBBARD, M. RICCHIUTO.*Discontinuous upwind residual distribution: a route to unconditional positivity and high order accuracy*, in "Comput. & Fluids", 2011, vol. 46, p. 263–269, http://dx.doi.org/10.1016/j.compfluid.2010.12.023.

[65] S. JIN.*Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations*, in "SIAM, J. Sci. Comput.", 1999, vol. 21, p. 441-454.

[66] R. JORDAN, D. KINDERLEHRER, F. OTTO.*The variational formulation of the Fokker-Planck equation*, in "SIAM J. Math. Anal.", 1998, vol. 29, n^{o} 1, p. 1–17.

[67] D. D. JOSEPH.*Fluid dynamics of two miscible liquids with diffusion and gradient stresses*, in "European J. Mech. B Fluids", 1990, vol. 9, n^{o} 6, p. 565–596.

[68] A. V. KAZHIKHOV, S. SMAGULOV.*The correctness of boundary value problems in a diffusion model in an inhomogeneous fluid*, in "Sov. Phys. Dokl.", 1977, vol. 22, p. 249–250.

[69] C. LIU, N. J. WALKINGTON.*Convergence of numerical approximations of the incompressible Navier-Stokes equations with variable density and viscosity*, in "SIAM J. Numer. Anal.", 2007, vol. 45, n^{o} 3, p. 1287–1304 (electronic), http://dx.doi.org/10.1137/050629008.

[70] P. MASON, A. AFTALION.*Classification of the ground states and topological defects in a rotating two-component Bose-Einstein condensate*, in "Phys. Rev. A", 2011, vol. 84, n^{o} 3, 033611.

[71] A. MELLET, A. VASSEUR.*On the barotropic compressible Navier-Stokes equations*, in "Comm. Partial Differential Equations", 2007, vol. 32, n^{o} 1-3, p. 431–452, http://dx.doi.org/10.1080/03605300600857079.

[72] F. OTTO.*The geometry of dissipative evolution equations: the porous medium equation*, in "Comm. Partial Differential Equations", 2001, vol. 26, n^{o} 1-2, p. 101–174.

[73] M. RICCHIUTO, R. ABGRALL.*Explicit Runge-Kutta residual distribution schemes for time dependent problems: second order case*, in "J. Comput. Phys.", 2010, vol. 229, n^{o} 16, p. 5653–5691, http://dx.doi.org/10.1016/j.jcp.2010.04.002.

[74] F. SANTAMBROGIO.*Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, Progress in Nonlinear Differential Equations and Their Applications 87, 1, Birkhäuser Basel, 2015, http://gen.lib.rus.ec/book/index.php?md5=24B4AA557102EC12148F101DF2C91937.

[75] C. VILLANI.*Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009, vol. 338, xxii+973, Old and new, http://dx.doi.org/10.1007/978-3-540-71050-9.

[76] M. VOHRALÍK.*Residual flux-based a posteriori error estimates for finite volume and related locally conservative methods*, in "Numer. Math.", 2008, vol. 111, n^{o} 1, p. 121–158, http://dx.doi.org/10.1007/s00211-008-0168-4.

[77] J. DE FRUTOS, B. GARCÍA-ARCHILLA, J. NOVO.*A posteriori error estimations for mixed finite-element approximations to the Navier-Stokes equations*, in "J. Comput. Appl. Math.", 2011, vol. 236, n^{o} 6, p. 1103–1122, http://dx.doi.org/10.1016/j.cam.2011.07.033.

Project-Team RMOD

Analyses and Languages Constructs for Object-Oriented Application Evolution

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Distributed programming and Software engineering

Table of contents

Project-Team RMOD

Creation of the Project-Team: 2009 July 01

Keywords:

Computer Science and Digital Science:

2. - Software
2.1. - Programming Languages
2.1.2. - Object-oriented programming
2.1.4. - Aspect-oriented programming
2.1.9. - Dynamic languages
2.1.10. - Domain-specific languages
2.5. - Software engineering
2.5.1. - Software Architecture & Design
2.5.3. - Empirical Software Engineering
2.5.4. - Software Maintenance & Evolution
2.6. - Infrastructure software
2.6.3. - Virtual machines

Other Research Topics and Application Domains:

2. - Health
2.7. - Medical devices
5. - Industry of the future
5.9. - Industrial maintenance
6.5. - Information systems
7. - Transport and logistics

1. Members

Research Scientists

Stéphane Ducasse [Team leader, Inria, Research Scientist, HDR]
Marcus Denker [Inria, Research Scientist]

Faculty Members

Nicolas Anquetil [Univ. Lille I, Faculty Member, HDR]
Damien Cassou [Univ. Lille I, Faculty Member, left for industry in Jul 2016]
Anne Etien [Univ. Lille I, Faculty Member, HDR]
Damien Pollet [Univ. Lille I, Faculty Member]

Technical Staff

Christophe Demarey [Inria, 70%]
Guillermo Polito [CNRS, 60%]
Olivier Auverlot [Univ. Lille I, 30%]
Pavel Krivanek [Inria, since Mar 2016]
Denis Kudriashov [Inria, since Oct 2015]
Esteban Lorenzano [Inria]
Nicolas Passerini [Inria, since Apr 2016]

PhD Students

Clément Béra [Inria, 2rd year, granted by Region Nord Pas de Calais]
Vincent Blondeau [Worldline, 2rd year, granted by CIFRE]
Gustavo Jansen de Souza Santos [Inria, 3rd year, granted by CNPq (Brazil)]
Brice Govin [Thales, 2rd year, granted by CIFRE]
Thibault Raffaillac [Inria, 2nd year, co-supervision with Stéphane Huot of Mjolnir team Inria]
Pablo Tesone [Inria-Ecole des Mines de Douai, 1st year]
Jason Lecerf [CEA-LIST, 1st year, co-supervision with Thierry Goubier]
Marco Naddeo [University of Torino, 3rd year, co-supervision with Pr. Viviana Bono]

Visiting Scientists

Klérisson Vinicius Ribeiro Da Paixão [University of Uberlandia (Brazil), 10 months internship from Oct 2015 to Jul 2016]
Abdelghani Alidra [Badji Mokhtar University (Algeria), 3 months internship from Nov 2015 to Jan 2016]
Cedrik Beler [ENIT, Apr 2016]
Alexandre Bergel [University of Chile]
Johan Fabry [University of Chile]
Ronie Salgado Faila [University of Chile, PhD Student, 6 months internship from Apr 2016 to Sep 2016]
Peter Uhnak [Czech Technical University in Prague, Master Student, from Apr 2016 to May 2016]

Administrative Assistant

Julie Jonas [Inria]

Others

Lionel Akue [Université de Lomé, Software Engineer, from Oct 2016 to Nov 2016]
Thibault Arloing [Univ. Lille I, Master student, from Apr 2016 until Sep 2016]
Julien Delplanque [Inria, Master Student, from Oct 2016 to Dec 2016]
Yann Dubois [Univ. Lille I, Master student, from Apr 2016 until Sep 2016]
Luc Fabresse [Ecole des Mines de Douai, Faculty Member, from Apr 2016]
Thomas Heniart [Univ. Lille I, Master student, until Jul 2016]
Skip Lentz [Inria, Master student, until Jan 2016]
Marion Noirbent [Univ. Lille I, Master student, from Apr 2016 until Aug 2016]
Maxime Roelandt [Univ. Lille I, Master student, from Apr 2016 until Sep 2016]
Valentin Ryckewaert [Univ. Lille I, Master student, from Apr 2016 until Aug 2016]

2. Overall Objectives

2.1. Introduction

Keywords: Software evolution, Maintenance, Program visualization, Program analyses, Meta modelling, Software metrics, Quality models, Object-oriented programming, Reflective programming, Traits, Dynamically typed languages, Dynamic Software Update, Pharo, Moose.

RMoD's general vision is defined in two objectives: remodularization and modularity constructs. These two views are the two faces of a same coin: maintenance could be eased with better engineering and analysis tools and programming language constructs could let programmers define more modular applications.

2.2. Reengineering and remodularization

While applications must evolve to meet new requirements, few approaches analyze the implications of their original structure (modules, packages, classes) and their transformation to support their evolution. Our research focuses on the *remodularization* of object-oriented applications. Automated approaches including clustering algorithms are not satisfactory because they often ignore user inputs. Our vision is that we need better approaches to support the transformation of existing software. The reengineering challenge tackled by RMoD is formulated as follows:

How to help remodularize existing software applications?

We are developing analyses and algorithms to remodularize object-oriented applications. This is why we started studying and building tools to support the *understanding of applications* at the level of packages and modules. This allows us to understand the results of the *analyses* that we are building.

We seek to create tools to help developers perform large refactoring. How can they keep track of changes in various locations in a system while ensuring *integrity of current and new code* by *uniformally applying new design choices*.

2.3. Constructs for modular and isolating programming languages

Dynamically-typed programming languages such as JavaScript are getting new attention as illustrated by the large investment of Google in the development of the Chrome V8 JavaScript engine and the development of a new dynamic language DART. This new trend is correlated to the increased adoption of dynamic programming languages for web-application development, as illustrated by Ruby on Rails, PHP and JavaScript. With web applications, users expect applications to be always available and getting updated on the fly. This continuous evolution of application is a real challenge [54]. Hot software evolution often requires *reflective* behavior and features. For instance in CLOS and Smalltalk each class modification automatically migrates existing instances on the fly.

At the same time, there is a need for *software isolationi.e.,*, applications should reliably run co-located with other applications in the same virtual machine with neither confidential information leaks nor vulnerabilities. Indeed, often for economical reasons, web servers run multiple applications on the same virtual machine. Users need confined applications. It is important that (1) an application does not access information of other applications running on the same virtual machine and (2) an application authorized to manipulate data cannot pass such authorization or information to other parts of the application that should not get access to it.

Static analysis tools have always been confronted to reflection [51]. Without a full treatment of reflection, static analysis tools are both incomplete and unsound. Incomplete because some parts of the program may not be included in the application call graph, and unsound because the static analysis does not take into account reflective features [60]. In reflective languages such as F-Script, Ruby, Python, Lua, JavaScript, Smalltalk and Java (to a certain extent), it is possible to nearly change any aspect of an application: change objects, change classes dynamically, migrate instances, and even load untrusted code.

Reflection and isolation concerns are a priori antagonistic, pulling language design in two opposite directions. Isolation, on the one hand, pulls towards more static elements and types (*e.g.,*, ownership types). Reflection, on the other hand, pulls towards fully dynamic behavior. This tension is what makes this a real challenge: As experts in reflective programming, dynamic languages and modular systems, we believe that by working on this important tension we can make a breakthrough and propose innovative solutions in resolving or mitigating this tension. With this endeavor, we believe that we are working on a key challenge that can have an impact on future programming languages. The language construct challenge tackled by RMoD is formulated as follows:

What are the language modularity constructs to support isolation?

In parallel we are continuing our research effort on traits [0] by assessing trait scalability and reuse on a large case study and developing a pure trait-based language. In addition, we dedicate efforts to remodularizing a meta-level architecture in the context of the design of an isolating dynamic language. Indeed at the extreme, modules and structural control of reflective features are the first steps towards flexible, dynamic, yet isolating, languages. As a result, we expect to demonstrate that having adequate composable units and scoping units will help the evolution and recomposition of an application.

[0]Traits are groups of methods that can be composed orthogonally to simple inheritance. Contrary to mixin, the class has the control of the composition and conflict management.

3. Research Program

3.1. Software Reengineering

Strong coupling among the parts of an application severely hampers its evolution. Therefore, it is crucial to answer the following questions: How to support the substitution of certain parts while limiting the impact on others? How to identify reusable parts? How to modularize an object-oriented application?

Having good classes does not imply a good application layering, absence of cycles between packages and reuse of well-identified parts. Which notion of cohesion makes sense in presence of late-binding and programming frameworks? Indeed, frameworks define a context that can be extended by subclassing or composition: in this case, packages can have a low cohesion without being a problem for evolution. How to obtain algorithms that can be used on real cases? Which criteria should be selected for a given remodularization?

To help us answer these questions, we work on enriching Moose, our reengineering environment, with a new set of analyses [45], [44]. We decompose our approach in three main and potentially overlapping steps:

1. Tools for understanding applications,
2. Remodularization analyses,
3. Software Quality.

3.1.1. Tools for understanding applications

Context and Problems. We are studying the problems raised by the understanding of applications at a larger level of granularity such as packages or modules. We want to develop a set of conceptual tools to support this understanding.

Some approaches based on Formal Concept Analysis (FCA) [75] show that such an analysis can be used to identify modules. However the presented examples are too small and not representative of real code.

Research Agenda.

FCA provides an important approach in software reengineering for software understanding, design anomalies detection and correction, but it suffers from two problems: (i) it produces lattices that must be interpreted by the user according to his/her understanding of the technique and different elements of the graph; and, (ii) the lattice can rapidly become so big that one is overwhelmed by the mass of information and possibilities [34]. We look for solutions to help people putting FCA to real use.

3.1.2. Remodularization analyses

Context and Problems. It is a well-known practice to layer applications with bottom layers being more stable than top layers [61]. Until now, few works have attempted to identify layers in practice: Mudpie [77] is a first cut at identifying cycles between packages as well as package groups potentially representing layers. DSM (dependency structure matrix) [76], [69] seems to be adapted for such a task but there is no serious empirical experience that validates this claim. From the side of remodularization algorithms, many were defined for procedural languages [57]. However, object-oriented programming languages bring some specific problems linked with late-binding and the fact that a package does not have to be systematically cohesive since it can be an extension of another one [78], [48].

As we are designing and evaluating algorithms and analyses to remodularize applications, we also need a way to understand and assess the results we are obtaining.

Research Agenda. We work on the following items:

Layer identification. We propose an approach to identify layers based on a semi-automatic classification of package and class interrelationships that they contain. However, taking into account the wish or knowledge of the designer or maintainer should be supported.

Cohesion Metric Assessment. We are building a validation framework for cohesion/coupling metrics to determine whether they actually measure what they promise to. We are also compiling a number of traditional metrics for cohesion and coupling quality metrics to evaluate their relevance in a software quality setting.

3.1.3. Software Quality

Research Agenda. Since software quality is fuzzy by definition and a lot of parameters should be taken into account we consider that defining precisely a unique notion of software quality is definitively a Grail in the realm of software engineering. The question is still relevant and important. We work on the two following items:

Quality models. We studied existing quality models and the different options to combine indicators — often, software quality models happily combine metrics, but at the price of losing the explicit relationships between the indicator contributions. There is a need to combine the results of one metric over all the software components of a system, and there is also the need to combine different metric results for any software component. Different combination methods are possible that can give very different results. It is therefore important to understand the characteristics of each method.

Bug prevention. Another aspect of software quality is validating or monitoring the source code to avoid the emergence of well known sources of errors and bugs. We work on how to best identify such common errors, by trying to identify earlier markers of possible errors, or by helping identifying common errors that programmers did in the past.

3.2. Language Constructs for Modular Design

While the previous axis focuses on how to help remodularizing existing software, this second research axis aims at providing new language constructs to build more flexible and recomposable software. We will build on our work on traits [73], [46] and classboxes [35] but also start to work on new areas such as isolation in dynamic languages. We will work on the following points: (1) Traits and (2) Modularization as a support for isolation.

3.2.1. Traits-based program reuse

Context and Problems. Inheritance is well-known and accepted as a mechanism for reuse in object-oriented languages. Unfortunately, due to the coarse granularity of inheritance, it may be difficult to decompose an application into an optimal class hierarchy that maximizes software reuse. Existing schemes based on single inheritance, multiple inheritance, or mixins, all pose numerous problems for reuse.

To overcome these problems, we designed a new composition mechanism called Traits [73], [46]. Traits are pure units of behavior that can be composed to form classes or other traits. The trait composition mechanism is an alternative to multiple or mixin inheritance in which the composer has full control over the trait composition. The result enables more reuse than single inheritance without introducing the drawbacks of multiple or mixin inheritance. Several extensions of the model have been proposed [43], [65], [36], [47] and several type systems were defined [49], [74], [66], [59].

Traits are reusable building blocks that can be explicitly composed to share methods across unrelated class hierarchies. In their original form, traits do not contain state and cannot express visibility control for methods. Two extensions, stateful traits and freezable traits, have been proposed to overcome these limitations. However, these extensions are complex both to use for software developers and to implement for language designers.

Research Agenda: Towards a pure trait language. We plan distinct actions: (1) a large application of traits, (2) assessment of the existing trait models and (3) bootstrapping a pure trait language.

- To evaluate the expressiveness of traits, some hierarchies were refactored, showing code reuse [38]. However, such large refactorings, while valuable, may not exhibit all possible composition problems, since the hierarchies were previously expressed using single inheritance and following certain patterns. We want to redesign from scratch the collection library of Smalltalk (or part of it). Such a redesign should on the one hand demonstrate the added value of traits on a real large and redesigned library and on the other hand foster new ideas for the bootstrapping of a pure trait-based language.

In particular we want to reconsider the different models proposed (stateless [46], stateful [37], and freezable [47]) and their operators. We will compare these models by (1) implementing a trait-based collection hierarchy, (2) analyzing several existing applications that exhibit the need for traits. Traits may be flattened [64]. This is a fundamental property that confers to traits their simplicity and expressiveness over Eiffel's multiple inheritance. Keeping these aspects is one of our priority in forthcoming enhancements of traits.

- Alternative trait models. This work revisits the problem of adding state and visibility control to traits. Rather than extending the original trait model with additional operations, we use a fundamentally different approach by allowing traits to be lexically nested within other modules. This enables traits to express (shared) state and visibility control by hiding variables or methods in their lexical scope. Although the traits' "flattening property" no longer holds when they can be lexically nested, the combination of traits with lexical nesting results in a simple and more expressive trait model. We formally specify the operational semantics of this combination. Lexically nested traits are fully implemented in AmbientTalk, where they are used among others in the development of a Morphic-like UI framework.
- We want to evaluate how inheritance can be replaced by traits to form a new object model. For this purpose we will design a minimal reflective kernel, inspired first from ObjVlisp [42] then from Smalltalk [52].

3.2.2. Reconciling Dynamic Languages and Isolation

Context and Problems. More and more applications require dynamic behavior such as modification of their own execution (often implemented using reflective features [56]). For example, F-script allows one to script Cocoa Mac-OS X applications and Lua is used in Adobe Photoshop. Now in addition more and more applications are updated on the fly, potentially loading untrusted or broken code, which may be problematic for the system if the application is not properly isolated. Bytecode checking and static code analysis are used to enable isolation, but such approaches do not really work in presence of dynamic languages and reflective features. Therefore there is a tension between the need for flexibility and isolation.

Research Agenda: Isolation in dynamic and reflective languages. To solve this tension, we will work on *Sure*, a language where isolation is provided by construction: as an example, if the language does not offer field access and its reflective facilities are controlled, then the possibility to access and modify private data is controlled. In this context, layering and modularizing the meta-level [39], as well as controlling the access to reflective features [40], [41] are important challenges. We plan to:

- Study the isolation abstractions available in erights (http://www.erights.org) [63], [62], and Java's class loader strategies [58], [53].
- Categorize the different reflective features of languages such as CLOS [55], Python and Smalltalk [67] and identify suitable isolation mechanisms and infrastructure [50].
- Assess different isolation models (access rights, capabilities [68]...) and identify the ones adapted to our context as well as different access and right propagation.
- Define a language based on
  - the decomposition and restructuring of the reflective features [39],
  - the use of encapsulation policies as a basis to restrict the interfaces of the controlled objects [72],
  - the definition of method modifiers to support controlling encapsulation in the context of dynamic languages.

An open question is whether, instead of providing restricted interfaces, we could use traits to grant additional behavior to specific instances: without trait application, the instances would only exhibit default public behavior, but with additional traits applied, the instances would get extra behavior. We will develop *Sure*, a modular extension of the reflective kernel of Smalltalk (since it is one of the languages offering the largest set of reflective features such as pointer swapping, class changing, class definition...) [67].

4. Application Domains

4.1. Programming Languages and Tools

Many of the results of RMoD are improving programming languages or development tools for such languages. As such the application domain of these results is as varied as the use of programming languages in general. Pharo, the language that RMoD develops, is used for a very broad range of applications. From pure research experiments to real world industrial use (the Pharo Consortium, http://consortium.pharo.org, has more than 20 company members) Examples are web applications, server backends for mobile applications or even graphical tools and embedded applications

4.2. Software Reengineering

Moose is a language-independent environment for reverse and re-engineering complex software systems. Moose provides a set of services including a common meta-model, metrics evaluation and visualization. As such Moose is used for analysing software systems to support understanding and continous development as well as software quality analysis.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Release of Pharo 5.0

We released a new version Pharo (Pharo 5.0) completly revisited with fundamental changes in the VM (object representation, compiler, ...)

5.1.2. HDR defenses

Anne Etien defended her HDR thesis.

5.1.3. Pharo web for the enterprise

A new book on Pharo.

5.1.4. Guillermo Polito hired as a CNRS engineer

Guillermo Polito a former PhD student in RMod was hired as a CNRS research engineer. This acknolewdges the quality of his research and work.

6. New Software and Platforms

6.1. Pharo

KEYWORDS: Live programmation objet - Reflective system
FUNCTIONAL DESCRIPTION

The platform Pharo is an open-source Smalltalk-inspired language and environment. It provides a platform for innovative development both in industry and research. By providing a stable and small core system, excellent developer tools, and maintained releases, Pharo's goal is to be a platform to build and deploy mission critical applications, while at the same time continue to evolve. In 2016, we released a new version Pharo (Pharo 5.0) completely revisited with fundamental changes in the VM (object representation, compiler, ...)

- Participants: Marcus Denker, Damien Cassou, Stephane Ducasse, Esteban Lorenzano, Damien Pollet, Igor Stasenko, Camillo Bruni, Camille Teruel and Clement Bera
- Partners: BetaNine - Debris publishing - École des Mines de Douai - HR Works - MAD - Pleiad - Sensus - Synectique - Université de Berne - Uqbar foundation Argentina - Vmware - Yesplan
- Contact: Marcus Denker
- URL: http://www.pharo.org

6.2. Moose

FUNCTIONAL DESCRIPTION

Moose is an extensive platform for software and data analysis. It offers multiple services ranging from importing and parsing data, to modeling, to measuring, querying, mining, and to building interactive and visual analysis tools.

- Participants: Stephane Ducasse, Muhammad Bhatti, Andre Cavalcante Hora, Nicolas Anquetil, Anne Etien, Guillaume Larcheveque and Alexandre Bergel
- Partners: Pleiad - Sensus - Synectique - Université de Berne - USI - Vrije Universiteit Brussel - Feenk
- Contact: Stephane Ducasse
- URL: http://www.moosetechnology.org

6.3. Pillar

KEYWORDS: HTML - LaTeX - HTML5

FUNCTIONAL DESCRIPTION

Pillar is a markup syntax and associated tools to write and generate documentation and books. Pillar is currently used to write several books and other documentation. Two platforms have already been created on top of Pillar: PillarHub and Marina.

- Contact: Damien Cassou
- URL: http://www.smalltalkhub.com/#!/~Pier/Pillar

7. New Results

7.1. Practical Validation of Bytecode to Bytecode JIT Compiler Dynamic Deoptimization.

Speculative inlining in just-in-time compilers enables many performance optimizations. However, it also introduces significant complexity. The compiler optimizations themselves, as well as the deoptimization mechanism are complex and error prone. To stabilize our bytecode to bytecode just-in-time compiler, we designed a new approach to validate the correctness of dynamic deoptimization. The approach consists of the symbolic execution of an optimized and an unop-timized bytecode compiled method side by side, deoptimizing the abstract stack at each deoptimization point (where dynamic deoptimization is possible) and comparing the deoptimized and unoptimized abstract stack to detect bugs. The implementation of our approach generated tests for several hundred thousands of methods, which are now available to be run automatically after each commit [13].

7.2. Recording and Replaying System-Specific Conventions.

In other situations, we found that developers sometimes perform sequences of code changes in a systematic way. These sequences consist of small code changes (*e.g.*, create a class, then extract a method to this class), which are applied to groups of related code entities (*e.g.*, some of the methods of a class). We propose to help this task by letting the developer record the sequence of code changes when he first applies it, and then generalize this sequence to apply it in other locations. The evaluation is based on real instances of such sequences that we found in different open source systems. We were able to replay 92% of the examples, which consisted in up to seven code entities modified up to 66 times. We are still working on the approach to allow for (semi-)automatic generalization of the recorded sequence of changes [71], [70].

7.3. Test Case Selection in Industry: an Analysis of Issues Related to Static Approaches

Automatic testing constitutes an important part of everyday development practice. But running all these tests may take hours. This is especially true for large systems involving, for example, the deployment of a web server or communication with a database. For this reason, tests are not launched as often as they should be and are mostly run at night. The company wishes to improve its development and testing process by giving to developers rapid feedback after a change. An interesting solution to give developers rapid feedback after a change is to reduce the number of tests to run by identifying only those exercising the piece of code changed. Two main approaches are proposed in the literature: static and dynamic. We evaluate these approaches on three industrial, closed source, cases to understand the strengths and weaknesses of each solution. We also propose a classification of problems that may arise when trying to identify the tests that cover a method.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

8.1.1. BlockChain

We started a new collaboration with a local company about tools and languages in the context of Blockchain systems. The collaboration started with a 2 month exploration phase involving an engineer at Inria Tech. A postdoc or PhD will start in 2017.

8.1.2. Worldline CIFRE

We are working on improving the testing behaviour of the developers. The PhD started in October 2014 and is ongoing.

8.1.3. Thales CIFRE

We are working on large industrial project rearchitecturization. The PhD started in January 2015 and is ongoing.

8.1.4. Pharo Consortium

The Pharo Consortium was founded in 2012 and is growing constantly. As of end 2016, it has 23 company members, 12 academic partners and 3 sponsoring companies. Inria supports the consortium with one full time engineer starting in 2011. More at http://consortium.pharo.org.

9. Partnerships and Cooperations

9.1. Regional Initiatives

Participants: Anne Etien, Nicolas Anquetil, Olivier Auverlot, Stéphane Ducasse. From Sept 2016 to Dec. 2018.

Lille Nord Europe European Associated Team with the PreCISE research center of Pr. A. Cleve from Namur University (Belgium).

This project aims to study the co-evolution between database structure and programs and to propose recommendations to perform required changes on cascade. These programs are either internal to the schema as functions or triggers or external as applications written in Java or Php built on top of the DB. Our intuition is that software engineering techniques can be efficient for such issues. This project also aims to unify the abstract representation of the DB and its relationships with the internal or external program.

9.2. European Initiatives

9.2.1. Collaborations with Major European Organizations

Marco Naddéo is a PhD student co-supervised by Damien Cassou, Stéphane Ducasse and Viviana Bono from University of Turin (Italy).

9.3. International Initiatives

9.3.1. Inria International Labs

Inria Chile
Associate Team involved in the International Lab:

9.3.1.1. PLOMO2

Title: Infrastructure for a new generation of development tools

International Partner (Institution - Laboratory - Researcher):

Universidad de Chile (Chile) - Computer Science Department, PLEIAD laboratory (DCC) - Alexander Bergel

Start year: 2014

See also: http://pleiad.cl/research/plomo2

Performing effective software development and maintenance are best achieved with effective tool support. Provided by a variety of tools, each one presenting a specific kinds of information supporting the task at hand. With Plomo2, we want to invent a new generation tools to navigate and profile programs by combining dynamic information with visualization to improve the development environment.

9.3.2. Inria Associate Teams Not Involved in an Inria International Labs

9.3.2.1. Informal International Partners

We are working with the Uqbar team from different argentinian universities. We hired three of the people: Nicolas Passerini(engineer), Esteban Lorenzano (engineer) and Pablo Tesone (PhD).

We are starting to work with Dr. Robert Pergl from the University of Prague.

9.3.3. Participation in Other International Programs

9.3.3.1. STIC AmSud projects

We were involved in two STIC AmSud projects:

Participants: Damien Cassou [correspondant], Gustavo Santos [RMoD], Martin Dias [RMoD], David Röthlisberger [UDP - Universidad Diego Portales, Santiago, Chile], Marcelo Almeida Maia [UFU - Federal University of Uberlândia, Brasil], Romain Robbes [Departamento de Ciencias de la Computación (DCC), Universidad de Chile, Santiago, Chile], Martin Monperrus [Spirals]. Project Partners: Inria RMOD, Inria Spirals, DCC Universidad de Chile, Universidad Diego Portale Chile, Federal University of Uberlândia, Brasil.

This project aims at facilitating the usage of frameworks and application programming interfaces (APIs) by mining software repositories. Our intuition is that mining reveals how existing projects instantiate these frameworks. By locating concrete framework instantiations in existing projects, we can recommend to developers the concrete procedures for how to use a particular framework for a particular task in a new system. Our project also tackles the challenge of adapting existing systems to new versions of a framework or API by seeking repositories for how other systems adapted to such changes. We plan to integrate recommendations of how to instantiate a framework and adapt to changes directly in the development environment. Those points taken together, considerably distinguish our approach from existing research in the area of framework engineering.

Thanks to this project, a PhD student of Federal University of Uberlândia in Brasil (Klérisson Vinícius Ribeiro da Paixão) did a six months internship in RMod, and prof. Marcelo Almeida Maia (from the same university) visited us for one week.

Participants: Nicolas Anquetil [correspondant], Anne Etien [RMoD], Gustavo Santos [RMoD], Marco Tulio Valente [UFMG - Federal University of Minas Gerais, Brazil], Alexander Bergel [Departamento de Ciencias de la Computación (DCC), Universidad de Chile, Santiago, Chile], Project Partners: Inria RMOD, DCC Universidad de Chile, Federal University of Minas Gerais, Brazil.

The goals of the collaboration is to provide tools to help software engineer restructure a large software system in an iterative and incremental way with input from both expert architect and advanced tool. The tools consist of: extraction of a model from source code, manipulation of the model to experiment with possible restructuring, architecture evaluation tool, recommendation tool to help the software engineers to define the new structure, and tool to back port the modification to the source code.

Pr. Marco Tulio Valente of Federal University of Minas Gerais (Brazil)visited RMod for one week.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Prof. Serge Demeyer, 2 days
- Prof. Marcelo Almeida Maia, 1 week, November 2016;
- Prof. Alain Plantec, 5 days, November 2016;
- Prof. Alexandre Bergel, 1 week, December 2016;
- Prof. Marco Tulio Valente, 1 week, December 2016;
- Eliot Miranda, Cadence,1 week, August 2016;
- Dr. Andrei Chis, 3 days, Decembrer 2016.

9.4.1.1. Internships

- Klérisson Vinícius Ribeiro da Paixão, Federal University of Uberlândia, Uberlândia (MG), Brazil, from September, 2015 to July, 2016;
- Lionel Akue, University of Lomé, Togo, from Oct 2016 to Nov 2016
- Julien Delplanque, University of Mons, Belgium, from September 2016 to December 2016.

9.4.2. Visits to International Teams

The whole RMod team spent two days in January at the Vrije Universiteit Brussel (Belgium) to discuss with the Software Languages Lab team: Coen De Roover, Elisa Gonzalez Boix, and their students.

Stéphane Ducasse and Damien Cassou visited Dr. Robert Pergl's team at Czech Technical University in Prague, one week in April.

9.4.2.1. Research Stays Abroad

Marcus Denker: Visit PLEIAD DCC University of Chile, Santiago de Chile 04/12-21/12. Visit in the context of the Inria Associated Team PLOMO2.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Selection

10.1.1.1. Chair of Conference Program Committees

- Anne Etien acted as Program chair for the following international events: IWST16 (International Worshop on SmallTalk Technologies) and VISSOFT 2016 (IEEE Working Conference on Software Visualization): NIER and Tool Track
- Stéphane Ducasse acted as Program Chair for ESUG 2016.

10.1.1.2. Member of the Conference Program Committees

- Stéphane Ducasse
  - International Conference 18th International Conference on Fundamental Approaches to Software Engineering (FASE) 2016.
  - For health reasons, I declined the PC participation to ICSE 2013, ECOOP 2013, ECOOP 2014, ECOOP 2015 (but I was the workshop chair) ECOOP 2016.
  - IWST'16 (International Worshop on SmallTalk Technologies).
- Anne Etien
  - DChanges'16 (4th International Workshop on Document Changes: Modeling, Detection, Storage and Visualization);
  - ME'16 (10th Workshop on Models and Evolution);
  - Modelwards'16 (4th International Conference on Model-Driven Engineering and Software Development).
  - IWST'16 (International Worshop on SmallTalk Technologies).
- Nicolas Anquetil
  - ICSOFT-EA'16 (11th International Conference on Software Technologies and Applications);
  - IWST'16 (International Worshop on SmallTalk Technologies).
- Marcus Denker
  - SANER 2017 (23rd IEEE International Conference on Software Analysis, Evolution, and Reengineering)
  - Meta 2016 (International Workshop, SPLASH 16)
  - SLE 2016 (International Conference on Software Language Engineering)
  - DLS 2016 (Dynamic Languages Symposium at SPLASH)
  - IWST'16 (International Worshop on SmallTalk Technologies).
- Clément Béra
  - ICOOOLPS'16 (11th Implementation, Compilation, Optimization of Object-Oriented Languages, Programs and Systems Workshop)
  - IWST'16 (International Worshop on SmallTalk Technologies).

10.1.2. Journal

10.1.2.1. Reviewer - Reviewing Activities

- Anne Etien reviewed manuscripts for the Software Quality Journal

- Nicolas Anquetil reviewed manuscripts for the ACM Transactions on Software Engineering and Methodology (TOSEM); Journal of Software: Evolution and Process (JSME); Computers & Operations Research (COR).
- Stéphane Ducasse reviewed manuscripts for Information ; Journal of Software: Evolution and Process (JSME); Journal of Science of Computer Programming; Journal of Object Technologies (JOT).
- Stéphane Ducasse and Marcus Denker are Moderators for arXiv cs.SE. http://arxiv.org/list/cs.SE/recent

10.1.3. Invited Talks

- Stéphane Ducasse gave a presentation at CITI laboratory Lyon, Diverse Team seminar Inria Rennes, La Sorbonne
- Clément Béra gave a presentation "Sista: Speculative inlining, Smalltalk-style", with Eliot Miranda (Cadence Design Systems) at the Stanford EE Computer Systems Colloquium, Stanford University (https://www.youtube.com/watch?v=f4Cvia-HZ-w).

10.1.4. Scientific Expertise

- Anne Etien reviewed several proposals for *Jeune Entreprise Innovante* and *Crédit Impôt Recherche*;
- Nicolas Anquetil acted as expert to evaluate project proposal for *Fondo Nacional de Desarrollo Cientifico y Tecnologico* (FONDECYT), Chile;
- Marcus Denker reviewed a research proposal for FONDECYT (Chilean National Science and Technology Commission);
- Damien Pollet reviewed projects proposal for NWO (Netherlands Organisation for Scientific Research) Veni grant proposal;
- Stéphane Ducasse participated to FWO (Belgium) PhD grant panel;
- Damien Cassou was a member of the Apps Togo jury, a contest from Ministry of Industry of Togo to encourage innovation in apps development (http://www.republicoftogo.com/Toutes-les-rubriques/Tech-Web/Voici-les-meilleures-applis-du-moment).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : C. Demarey, Intégration Continue, 4h, M2, GIS2A5, Polytech-Lille, France

Master : C. Demarey, Intégration Continue, 12h, M2, IAGL, Université de Lille 1, France

Master: Anne Etien, Metamodelisation for Reverse Engineering, 25h, M2, Montpellier, France

Master: Anne Etien, Test et Maintenance, 10h, M2, Polytech-Lille, France

Master: Anne Etien, Test et Maintenance, 14h, M2, Polytech-Lille, France

Master : Anne Etien, Système d'information objet, 22h, M1, Polytech-Lille, France

Master : Anne Etien, Bases de données, 44h, M1, Polytech-Lille, France

Master : Anne Etien, Bases de données Avancées, 12h, M1, Polytech-Lille, France

Licence : Anne Etien, Bases de données, 22h, L3, Polytech-Lille, France

Licence : Nicolas Anquetil, Conception et programmation objet avancées, 52h, L2, Univ. Lille1, IUT-A, France

Licence : Nicolas Anquetil, Principes des systèmes d'exploitation, 40h, L2, Univ. Lille1, IUT-A, France

Licence : Nicolas Anquetil, Modélisation mathématique, 14h, L2, univ. Lille1, IUT-A, France

Licence : Nicolas Anquetil, Conception et développement d'applications mobiles, 30h, L2, univ. Lille1, IUT-A, France

Licence : Nicolas Anquetil, Méthodologie de la production d'applications, 33h, L2, univ. Lille1, IUT-A, France

Licence : Damien Pollet, Java lecture, 30h, L3, Telecom Lille, France

Licence : Damien Pollet, Java lecture, 20h, L3, Telecom Lille, France

Licence : Damien Pollet, Java project, 20h, L3, Telecom Lille, France

Licence : Damien Pollet, Computer architecture,10h, L3, Telecom Lille, France

Master : Damien Pollet, Technologies for information systems, 30h, M1, Telecom Lille, France

Master : Damien Pollet, Network algorithms, 30h, M2, Telecom Lille, France

Master : Damien Pollet, Software engineering, 6h, M1, Telecom Lille, France

Licence: Damien Cassou, Algorithmique et programmation, 25h, L1, Polytech-Lille, France

Licence: Damien Cassou, OpenDevs—Communauté de Développeurs, 30h, L3, Univ. Lille-1, France

Licence : Clément Béra, Algorithmique et programmation, 35h, L1, Polytech-Lille, France

Licence : Clément Béra, Conception orienté objet, 35h, L1, Univ. Lille-1, France

Master : Clément Béra, Génie Logiciel, 40h, M3, Univ. Lille-1, France

Master : Vincent Blondeau, Bases de données, 12h, M1, Polytech-Lille, France

E-learning

Mooc: Stéphane Ducasse, Damien Cassou, Luc Fabresse, "Programmation Objet Immersive en Pharo / Live Object Programming in Pharo", 7 weeks, by France Université Numerique and Inria, for programmers, 3400 participants

10.2.2. Supervision

PhD in progress : Vincent Blondeau, "Studying and Enhancing Testing Habits of Developers in a Large IT Company", CIFRE Worldline, started Oct 2014, Anne Etien, Nicolas Anquetil.

PhD in progress : Brice Govin, "Support to implement a rejuvenated software architecture in legacy software", CIFRE Thales, started Jan 2015, Anne Etien, Nicolas Anquetil, Stéphane Ducasse.

PhD in progress : Gustavo Santos, "Assessing and Improving Code Transformations to Support Software Evolution", started April 2014, Anne Etien, Nicolas Anquetil.

PhD: Camille Teruel, "Security for dynamic languages", 21/01/2016, Stéphane Ducasse, Damien Cassou

PhD in progress: Marco Naddéo, "Evolvable and reusable software: from language design to developer tools", started 01/01/2014, Viviana Bono (University of Turin), Damien Cassou, Stéphane Ducasse

PhD in progress: Clément Béra, "Evolvable Runtimes for a Changing World", started Oct 2014,Stéphane Ducasse, Marcus Denker.

PhD in progresss: Klérisson Paixao, "Automating reactions for API changes based on dynamic analysis", started 16/08/2014, Marcelo Maia, Damien Cassou, Nicolas Anquetil.

10.2.3. Juries

- Anne Etien: Jonathan Pepin, PhD thesis, "Architecture d'entreprise : alignement des cartographies me´tiers et applicatives du syste`me d'informations", Dec. 5th, 2016, University of Nantes (France).
- Nicolas Anquetil: David Michael Cutting, PhD thesis, "Enhancing Legacy Software System Analysis by Combining Behavioural and Semantic Information Services", Aug. 5th, 2016, University of East Anglia, Norwich (England).

10.3. Popularization

- Stéphane Ducasse and Damien Cassou were important contributors of the “Enterprise Pharo” book (http://files.pharo.org/books/enterprise-pharo/).
- Brice Govin and Vincent Blondeau participated in the contest “My thesis in 180 seconds”

11. Bibliography

Major publications by the team in recent years

[1] N. Anquetil, K. M. de Oliveira, K. D. de Sousa, M. G. Batista Dias. *Software maintenance seen as a knowledge management issue*, in "Information Software Technology", 2007, vol. 49, n^{o} 5, p. 515–529 [*DOI :* 10.1016/j.infsof.2006.07.007], http://rmod.lille.inria.fr/archives/papers/Anqu07a-IST-MaintenanceKnowledge.pdf.

[2] M. Denker, S. Ducasse, É. Tanter. *Runtime Bytecode Transformation for Smalltalk*, in "Journal of Computer Languages, Systems and Structures", July 2006, vol. 32, n^{o} 2-3, p. 125–139 [*DOI :* 10.1016/j.cl.2005.10.002], http://rmod.lille.inria.fr/archives/papers/Denk06a-COMLAN-RuntimeByteCode.pdf.

[3] S. Ducasse, O. Nierstrasz, N. Schärli, R. Wuyts, A. P. Black. *Traits: A Mechanism for fine-grained Reuse*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", March 2006, vol. 28, n^{o} 2, p. 331–388 [*DOI :* 10.1145/1119479.1119483], http://scg.unibe.ch/archive/papers/Duca06bTOPLASTraits.pdf.

[4] S. Ducasse, D. Pollet. *Software Architecture Reconstruction: A Process-Oriented Taxonomy*, in "IEEE Transactions on Software Engineering", July 2009, vol. 35, n^{o} 4, p. 573-591 [*DOI :* 10.1109/TSE.2009.19], http://rmod.lille.inria.fr/archives/papers/Duca09c-TSE-SOAArchitectureExtraction.pdf.

[5] S. Ducasse, D. Pollet, M. Suen, H. Abdeen, I. Alloui. *Package Surface Blueprints: Visually Supporting the Understanding of Package Relationships*, in "ICSM'07: Proceedings of the IEEE International Conference on Software Maintenance", 2007, p. 94–103, http://scg.unibe.ch/archive/papers/Duca07cPackageBlueprintICSM2007.pdf.

[6] A. Kuhn, S. Ducasse, T. Gîrba. *Semantic Clustering: Identifying Topics in Source Code*, in "Information and Software Technology", March 2007, vol. 49, n^{o} 3, p. 230–243 [*DOI :* 10.1016/j.infsof.2006.10.017], http://scg.unibe.ch/archive/drafts/Kuhn06bSemanticClustering.pdf.

[7] J. Laval, S. Denier, S. Ducasse, A. Bergel. *Identifying cycle causes with Enriched Dependency Structural Matrix*, in "WCRE '09: Proceedings of the 2009 16th Working Conference on Reverse Engineering", Lille, France, 2009, http://rmod.lille.inria.fr/archives/papers/Lava09c-WCRE2009-eDSM.pdf.

[8] O. Nierstrasz, S. Ducasse, T. Gîrba. *The Story of Moose: an Agile Reengineering Environment*, in "Proceedings of the European Software Engineering Conference", New York NY, M. Wermelinger, H. Gall (editors), ESEC/FSE'05, ACM Press, 2005, p. 1–10, Invited paper [*DOI :* 10.1145/1095430.1081707], http://scg.unibe.ch/archive/papers/Nier05cStoryOfMoose.pdf.

[9] J. SINGER, T. LETHBRIDGE, N. VINSON, N. ANQUETIL.*An examination of software engineering work practices*, in "Proceedings of the 1997 conference of the Centre for Advanced Studies on Collaborative research", CASCON '97, IBM Press, 1997, http://rmod.lille.inria.fr/archives/papers/Sing97a-SoftwareEngineeringWorkPractices.pdf.

[10] S. C. B. DE SOUZA, N. ANQUETIL, K. M. DE OLIVEIRA.*A study of the documentation essential to software maintenance*, in "Proceedings of the 23rd annual international conference on Design of communication: documenting & designing for pervasive information", New York, NY, USA, SIGDOC '05, ACM, 2005, p. 68–75, http://dx.doi.org/10.1145/1085313.1085331.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. ETIEN.*Metamodelisation to support Test and Evolution*, Université de Lille, June 2016, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01352817.

[12] C. TERUEL.*Adaptability and Encapsulation in Dynamically-Typed Languages: Taming Reflection and Extension Methods*, Université des Sciences et Technologies de Lille, January 2016, https://tel.archives-ouvertes.fr/tel-01290364.

Articles in International Peer-Reviewed Journal

[13] C. BERA, E. MIRANDA, M. DENKER, S. DUCASSE.*Practical Validation of Bytecode to Bytecode JIT Compiler Dynamic Deoptimization*, in "Journal of Object Technology (JOT)", April 2016, vol. 15, n^{o} 2, p. 1:1-26 [*DOI :* 10.5381/JOT.2016.15.2.A1], https://hal.inria.fr/hal-01299371.

[14] V. BLONDEAU, A. ETIEN, N. ANQUETIL, S. CRESSON, P. CROISY, S. DUCASSE.*Test Case Selection in Industry: an Analysis of Issues Related to Static Approaches*, in "Software Quality Journal", 2016 [*DOI :* 10.1007/S11219-016-9328-4], https://hal.inria.fr/hal-01344842.

[15] J. DE KOSTER, S. MARR, T. VAN CUTSEM, T. D 'HONDT.*Domains: Sharing State in the Communicating Event-Loop Actor Model*, in "Computer Languages, Systems and Structures", January 2016 [*DOI :* 10.1016/J.CL.2016.01.003], https://hal.inria.fr/hal-01273665.

[16] M. DE WAEL, S. MARR, B. DE FRAINE, T. VAN CUTSEM, W. DE MEUTER.*Partitioned Global Address Space Languages*, in "ACM Computing Surveys", January 2016, 29, https://hal.inria.fr/hal-01109405.

[17] A. HORA, R. ROBBES, M. TULIO VALENTE, N. ANQUETIL, A. ETIEN, S. DUCASSE.*How do Developers React to API Evolution? a Large-Scale Empirical Study*, in "Software Quality Journal", October 2016 [*DOI :* 10.1007/S11219-016-9344-4], https://hal.inria.fr/hal-01417930.

[18] C. MAFFORT, M. T. VALENTE, R. TERRA, M. BIGONHA, N. ANQUETIL, A. HORA.*Mining Architectural Violations from Version History*, in "Empirical Software Engineering", June 2016, vol. 21, n^{o} 3, 41 [*DOI :* 10.1007/S10664-014-9348-2], https://hal.inria.fr/hal-01075642.

[19] M. SMID, F. G. RODRÍGUEZ-GONZÁLEZ, A. M. SIEUWERTS, R. SALGADO, W. J. C. PRAGER-VAN DER SMISSEN, M. V. D. VLUGT-DAANE, A. VAN GALEN, S. NIK-ZAINAL, J. STAAF, A. B. BRINKMAN, M. J. VAN DE VIJVER, A. L. RICHARDSON, A. FATIMA, K. BERENTSEN, A. BUTLER, S. MARTIN, H. R. DAVIES, R. DEBETS, M. E. M.-V. GELDER, C. H. M. VAN DEURZEN, G. MACGROGAN, G. G. G. M. VAN

DEN EYNDEN, C. PURDIE, A. M. THOMPSON, C. CALDAS, P. N. SPAN, P. T. SIMPSON, S. R. LAKHANI, S. VAN LAERE, C. DESMEDT, M. RINGNÉR, S. TOMMASI, J. EYFORD, A. BROEKS, A. VINCENT-SALOMON, P. A. FUTREAL, S. KNAPPSKOG, T. A. KING, G. THOMAS, A. VIARI, A. LANGERØD, A.-L. BØRRESEN-DALE, E. BIRNEY, H. G. STUNNENBERG, M. STRATTON, J. A. FOEKENS, J. W. M. MARTENS.*Breast cancer genome and transcriptome integration implicates specific mutational signatures with immune cell infiltration*, in "Nature Communications", 2016, vol. 7 [*DOI :* 10.1038/NCOMMS12910], https://hal.inria.fr/hal-01388445.

International Conferences with Proceedings

[20] A. CAVALCANTE HORA, M. TULIO VALENTE, R. ROBBES, N. ANQUETIL.*When should internal interfaces be promoted to public?*, in "FSE 2016 Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering", Seattle, United States, November 2016 [*DOI :* 10.1145/2950290.2950306], https://hal.inria.fr/hal-01417888.

[21] J. P. SANDOVAL ALCOCER, M. DENKER, A. BERGEL, Y. ACURANA.*Dynamically Composing Collection Operations through Collection Promises*, in "Proceedings of the 11th Edition of the International Workshop on Smalltalk Technologies", Prague, Czech Republic, Proceedings of the 11th Edition of the International Workshop on Smalltalk Technologies, ESUG, August 2016, p. 8:1–8:5 [*DOI :* 10.1145/2991041.2991049], https://hal.inria.fr/hal-01358347.

Conferences without Proceedings

[22] T. ARLOING, Y. DUBOIS, D. CASSOU, S. DUCASSE.*Pillar: A Versatile and Extensible Lightweight Markup Language*, in "International Workshop on Smalltalk Technologies", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01353882.

[23] C. BERA.*A low Overhead Per Object Write Barrier for the Cog VM*, in "IWST 16", Pragues, Czech Republic, 2016, https://hal.archives-ouvertes.fr/hal-01356338.

[24] V. BLONDEAU, N. ANQUETIL, S. DUCASSE, S. CRESSON, P. CROISY.*Test Selection with Moose In Industry*, in "IWST'16", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01352468.

[25] S. DUCASSE, E. MIRANDA, A. PLANTEC.*Pragmas: Literal Messages as Powerful Method Annotations*, in "International Workshop on Smalltalk Technologies - IWST 2016", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01353592.

[26] B. GOVIN, A. MONEGIER DU SORBIER, N. ANQUETIL, S. DUCASSE.*Clustering technique for conceptual clusters*, in "IWST'16 International Workshop on Smalltalk Technologies", Prague, Czech Republic, August 2016, https://hal.archives-ouvertes.fr/hal-01353205.

[27] N. MILOJKOVIĆ, C. BERA, M. GHAFARI, O. NIERSTRASZ.*Inferring Types by Mining Class Usage Frequency from Inline Caches*, in "IWST'16", Pragues, Czech Republic, 2016, https://hal.archives-ouvertes.fr/hal-01357071.

[28] M. RIZUN, G. SANTOS, S. DUCASSE, C. TERUEL.*Phorms: Pattern Combinator Library for Pharo*, in "International Workshop on Smalltalk Technologies", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01353883.

[29] R. SALGADO, S. DUCASSE.*Lowcode: Extending Pharo with C Types to Improve Performance*, in "International Workshop on Smalltalk Technologies", Prague, Czech Republic, August 2016, https://hal.inria.fr/hal-01353884.

Research Reports

[30] A. BERGEL, S. DUCASSE, M. DENKER, J. FABRY.*PLOMO2 Associate Team Final Report*, Inria, October 2016, https://hal.inria.fr/hal-01389983.

[31] M. DENKER, N. ANQUETIL, D. CASSOU, S. DUCASSE, A. ETIEN, D. POLLET.*Project-Team RMoD 2015 Activity Report*, Inria Lille - Nord Europe, February 2016, https://hal.inria.fr/hal-01267026.

Scientific Popularization

[32] O. AUVERLOT, S. DUCASSE.*Un Chat en Pharo*, January 2016, Article in GNU/Linux Magazine, two parts http://connect.ed-diamond.com/GNU-Linux-Magazine/GLMF-189/Un-chat-en-Pharo-le-serveur http://connect.ed-diamond.com/GNU-Linux-Magazine/GLMF-189/Un-chat-en-Pharo-le-client2, https://hal.inria.fr/hal-01353594.

Other Publications

[33] N. PAPOULIAS, M. DENKER, S. DUCASSE, L. FABRESSE.*End-User Abstractions for Meta-Control: Reifying the Reflectogram*, January 2017, working paper or preprint, https://hal.inria.fr/hal-01424787.

References in notes

[34] N. ANQUETIL.*A Comparison of Graphs of Concept for Reverse Engineering*, in "Proceedings of the 8th International Workshop on Program Comprehension", Washington, DC, USA, IWPC'00, IEEE Computer Society, 2000, p. 231–, http://rmod.lille.inria.fr/archives/papers/Anqu00b-ICSM-GraphsConcepts.pdf.

[35] A. BERGEL, S. DUCASSE, O. NIERSTRASZ.*Classbox/J: Controlling the Scope of Change in Java*, in "Proceedings of 20th International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA'05)", New York, NY, USA, ACM Press, 2005, p. 177–189 [*DOI :* 10.1145/1094811.1094826], http://scg.unibe.ch/archive/papers/Berg05bclassboxjOOPSLA.pdf.

[36] A. BERGEL, S. DUCASSE, O. NIERSTRASZ, R. WUYTS.*Stateful Traits*, in "Advances in Smalltalk — Proceedings of 14th International Smalltalk Conference (ISC 2006)", LNCS, Springer, August 2007, vol. 4406, p. 66–90, http://dx.doi.org/10.1007/978-3-540-71836-9_3.

[37] A. BERGEL, S. DUCASSE, O. NIERSTRASZ, R. WUYTS.*Stateful Traits and their Formalization*, in "Journal of Computer Languages, Systems and Structures", 2008, vol. 34, n^o 2-3, p. 83–108, http://dx.doi.org/10.1016/j.cl.2007.05.003.

[38] A. P. BLACK, N. SCHÄRLI, S. DUCASSE.*Applying Traits to the Smalltalk Collection Hierarchy*, in "Proceedings of 17th International Conference on Object-Oriented Programming Systems, Languages and Applications (OOPSLA'03)", October 2003, vol. 38, p. 47–64 [*DOI :* 10.1145/949305.949311], http://scg.unibe.ch/archive/papers/Blac03aTraitsHierarchy.pdf.

[39] G. BRACHA, D. UNGAR.*Mirrors: design principles for meta-level facilities of object-oriented programming languages*, in "Proceedings of the International Conference on Object-Oriented Programming, Systems,

Languages, and Applications (OOPSLA'04), ACM SIGPLAN Notices", New York, NY, USA, ACM Press, 2004, p. 331–344, http://bracha.org/mirrors.pdf.

[40] D. CAROMEL, J. VAYSSIÈRE.*Reflections on MOPs, Components, and Java Security*, in "ECOOP '01: Proceedings of the 15th European Conference on Object-Oriented Programming", Springer-Verlag, 2001, p. 256–274.

[41] D. CAROMEL, J. VAYSSIÈRE.*A security framework for reflective Java applications*, in "Software: Practice and Experience", 2003, vol. 33, n^O 9, p. 821–846, http://dx.doi.org/10.1002/spe.528.

[42] P. COINTE.*Metaclasses are First Class: the ObjVlisp Model*, in "Proceedings OOPSLA '87, ACM SIGPLAN Notices", December 1987, vol. 22, p. 156–167.

[43] S. DENIER.*Traits Programming with AspectJ*, in "Actes de la Première Journée Francophone sur le Développement du Logiciel par Aspects (JFDLPA'04)", Paris, France, P. COINTE (editor), September 2004, p. 62–78.

[44] S. DUCASSE, T. GÎRBA.*Using Smalltalk as a Reflective Executable Meta-Language*, in "International Conference on Model Driven Engineering Languages and Systems (Models/UML 2006)", Berlin, Germany, LNCS, Springer-Verlag, 2006, vol. 4199, p. 604–618 [*DOI :* 10.1007/11880240_42], http://scg.unibe.ch/archive/papers/Duca06dMOOSEMODELS2006.pdf.

[45] S. DUCASSE, T. GÎRBA, M. LANZA, S. DEMEYER.*Moose: a Collaborative and Extensible Reengineering Environment*, in "Tools for Software Maintenance and Reengineering", Milano, RCOST / Software Technology Series, Franco Angeli, 2005, p. 55–71, http://scg.unibe.ch/archive/papers/Duca05aMooseBookChapter.pdf.

[46] S. DUCASSE, O. NIERSTRASZ, N. SCHÄRLI, R. WUYTS, A. P. BLACK.*Traits: A Mechanism for fine-grained Reuse*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", March 2006, vol. 28, n^O 2, p. 331–388 [*DOI :* 10.1145/1119479.1119483], http://scg.unibe.ch/archive/papers/Duca06bTOPLASTraits.pdf.

[47] S. DUCASSE, R. WUYTS, A. BERGEL, O. NIERSTRASZ.*User-Changeable Visibility: Resolving Unanticipated Name Clashes in Traits*, in "Proceedings of 22nd International Conference on Object-Oriented Programming, Systems, Languages, and Applications (OOPSLA'07)", New York, NY, USA, ACM Press, October 2007, p. 171–190 [*DOI :* 10.1145/1297027.1297040], http://scg.unibe.ch/archive/papers/Duca07b-FreezableTrait.pdf.

[48] A. DUNSMORE, M. ROPER, M. WOOD.*Object-Oriented Inspection in the Face of Delocalisation*, in "Proceedings of ICSE '00 (22nd International Conference on Software Engineering)", ACM Press, 2000, p. 467–476.

[49] K. FISHER, J. REPPY.*Statically typed traits*, University of Chicago, Department of Computer Science, December 2003, n^O TR-2003-13, http://www.cs.uchicago.edu/research/publications/techreports/TR-2003-13.

[50] P. W. L. FONG, C. ZHANG.*Capabilities as alias control: Secure cooperation in dynamically extensible systems*, Department of Computer Science, University of Regina, 2004.

[51] M. FURR, J.-H. AN, J. S. FOSTER.*Profile-guided static typing for dynamic scripting languages*, in "OOPSLA'09", 2009.

[52] A. GOLDBERG.*Smalltalk 80: the Interactive Programming Environment*, Addison Wesley, Reading, Mass., 1984.

[53] L. GONG.*New security architectural directions for Java*, in "compcon", 1997, vol. 0, 97, http://dx.doi.org/10.1109/CMPCON.1997.584679.

[54] M. HICKS, S. NETTLES.*Dynamic software updating*, in "ACM Transactions on Programming Languages and Systems", nov 2005, vol. 27, n^{o} 6, p. 1049–1096, http://dx.doi.org/10.1145/1108970.1108971.

[55] G. KICZALES, J. DES RIVIÈRES, D. G. BOBROW.*The Art of the Metaobject Protocol*, MIT Press, 1991.

[56] G. KICZALES, L. RODRIGUEZ.*Efficient Method Dispatch in PCL*, in "Proceedings of ACM conference on Lisp and Functional Programming", Nice, 1990, p. 99–105.

[57] R. KOSCHKE.*Atomic Architectural Component Recovery for Program Understanding and Evolution*, Universität Stuttgart, 2000, http://www2.informatik.uni-stuttgart.de/cgi-bin/NCSTRL/NCSTRL_view.pl?id=DIS-2000-05&mod=0&engl=0&inst=PS.

[58] S. LIANG, G. BRACHA.*Dynamic Class Loading in the Java Virtual Machine*, in "Proceedings of OOPSLA '98, ACM SIGPLAN Notices", 1998, p. 36–44.

[59] L. LIQUORI, A. SPIWACK.*FeatherTrait: A Modest Extension of Featherweight Java*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", 2008, vol. 30, n^{o} 2, p. 1–32 [*DOI :* 10.1145/1330017.1330022], http://www-sop.inria.fr/members/Luigi.Liquori/PAPERS/toplas-07.pdf.

[60] B. LIVSHITS, T. ZIMMERMANN.*DynaMine: finding common error patterns by mining software revision histories*, in "SIGSOFT Software Engineering Notes", September 2005, vol. 30, n^{o} 5, p. 296-305.

[61] R. C. MARTIN.*Agile Software Development. Principles, Patterns, and Practices*, Prentice-Hall, 2002.

[62] M. S. MILLER.*Robust Composition: Towards a Unified Approach to Access Control and Concurrency Control*, Johns Hopkins University, Baltimore, Maryland, USA, May 2006.

[63] M. S. MILLER, C. MORNINGSTAR, B. FRANTZ.*Capability-based Financial Instruments*, in "FC '00: Proceedings of the 4th International Conference on Financial Cryptography", Springer-Verlag, 2001, vol. 1962, p. 349–378.

[64] O. NIERSTRASZ, S. DUCASSE, N. SCHÄRLI.*Flattening Traits*, in "Journal of Object Technology", May 2006, vol. 5, n^{o} 4, p. 129–148, http://www.jot.fm/issues/issue_2006_05/article4.

[65] P. J. QUITSLUND.*Java Traits — Improving Opportunities for Reuse*, OGI School of Science & Engineering, Beaverton, Oregon, USA, September 2004, n^{o} CSE-04-005.

[66] J. REPPY, A. TURON.*A Foundation for Trait-based Metaprogramming*, in "International Workshop on Foundations and Developments of Object-Oriented Languages", 2006.

[67] F. RIVARD.*Pour un lien d'instanciation dynamique dans les langages à classes*, in "JFLA96", Inria — collection didactique, January 1996.

[68] J. H. SALTZER, M. D. SCHOROEDER.*The Protection of Information in Computer Systems*, in "Fourth ACM Symposium on Operating System Principles", IEEE, September 1975, vol. 63, p. 1278–1308.

[69] N. SANGAL, E. JORDAN, V. SINHA, D. JACKSON.*Using Dependency Models to Manage Complex Software Architecture*, in "Proceedings of OOPSLA'05", 2005, p. 167–176.

[70] G. SANTOS, N. ANQUETIL, A. ETIEN, S. DUCASSE, M. TULIO VALENTE.*System Specific, Source Code Transformations*, in "31st IEEE International Conference on Software Maintenance and Evolution (ICSME)", Bremen, Germany, September 2015, 10, https://hal.inria.fr/hal-01185637.

[71] G. SANTOS, A. ETIEN, N. ANQUETIL, S. DUCASSE, M. TULIO VALENTE.*Recording and Replaying System Specific, Source Code Transformations*, in "15th IEEE International Working Conference on Source Code Analysis and Manipulation (SCAM)", Bremen, Germany, September 2015, 10, https://hal.inria.fr/hal-01185639.

[72] N. SCHÄRLI, A. P. BLACK, S. DUCASSE.*Object-oriented Encapsulation for Dynamically Typed Languages*, in "Proceedings of 18th International Conference on Object-Oriented Programming Systems, Languages and Applications (OOPSLA'04)", October 2004, p. 130–149 [*DOI :* 10.1145/1028976.1028988], http://scg.unibe.ch/archive/papers/Scha04bOOEncapsulation.pdf.

[73] N. SCHÄRLI, S. DUCASSE, O. NIERSTRASZ, A. P. BLACK.*Traits: Composable Units of Behavior*, in "Proceedings of European Conference on Object-Oriented Programming (ECOOP'03)", LNCS, Springer Verlag, July 2003, vol. 2743, p. 248–274 [*DOI :* 10.1007/B11832], http://scg.unibe.ch/archive/papers/Scha03aTraits.pdf.

[74] C. SMITH, S. DROSSOPOULOU.*Chai: Typed Traits in Java*, in "Proceedings ECOOP 2005", 2005.

[75] G. SNELTING, F. TIP.*Reengineering Class Hierarchies using Concept Analysis*, in "ACM Trans. Programming Languages and Systems", 1998.

[76] K. J. SULLIVAN, W. G. GRISWOLD, Y. CAI, B. HALLEN.*The Structure and Value of Modularity in SOftware Design*, in "ESEC/FSE 2001", 2001.

[77] D. VAINSENCHER.*MudPie: layers in the ball of mud*, in "Computer Languages, Systems & Structures", 2004, vol. 30, n^{o} 1-2, p. 5–19.

[78] N. WILDE, R. HUITT.*Maintenance Support for Object-Oriented Programs*, in "IEEE Transactions on Software Engineering", December 1992, vol. SE-18, n^{o} 12, p. 1038–1044.

Project-Team SEQUEL

Sequential Learning

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
Université Charles de Gaulle (Lille 3)

Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Optimization, machine learning and statistical methods

Table of contents

Project-Team SEQUEL

Creation of the Project-Team: 2007 July 01

Keywords:

Computer Science and Digital Science:

3. - Data and knowledge
3.1. - Data
3.1.1. - Modeling, representation
3.1.4. - Uncertain data
3.3. - Data and knowledge analysis
3.3.1. - On-line analytical processing
3.3.2. - Data mining
3.3.3. - Big data analysis
3.4. - Machine learning and statistics
3.4.1. - Supervised learning
3.4.2. - Unsupervised learning
3.4.3. - Reinforcement learning
3.4.4. - Optimization and learning
3.4.6. - Neural networks
3.4.8. - Deep learning
3.5.2. - Recommendation systems
4.8. - Privacy-enhancing technologies
5.1. - Human-Computer Interaction
8. - Artificial intelligence
8.2. - Machine learning
8.3. - Signal analysis
8.7. - AI algorithmics

Other Research Topics and Application Domains:

5.8. - Learning and training
6.1. - Software industry
6.1.1. - Software engineering
6.1.2. - Software evolution, maintenance
9.1.1. - E-learning, MOOC
9.4. - Sciences
9.4.5. - Data science

1. Members

Research Scientists

Emilie Kaufmann [CNRS, Researcher]
Alessandro Lazaric [Inria, Researcher]
Odalric Maillard [Inria, Researcher, moved to SEQUEL on 11/1/2016]

Rémi Munos [secondment at Google/Deepmind, Senior Researcher, HDR]
Daniil Ryabko [Inria, Researcher, HDR]
Michal Valko [Inria, Researcher, HDR]

Faculty Members

Philippe Preux [Team leader, Univ. Lille III, Professor, HDR]
Christos Dimitrakakis [Univ. Lille III, Associate Professor, HDR]
Romaric Gaudel [Univ. Lille III, Associate Professor]
Jérémie Mary [Univ. Lille III, Associate Professor, HDR]
Bilal Piot [Univ. Lille I, Associate Professor]
Olivier Pietquin [Univ. Lille I, Professor, currently in secondment at Google/Deepmind since May 2016, HDR]

PhD Students

Marc Abeille [Univ. Lille I]
Merwan Barlier [Orange Labs, granted by CIFRE]
Alexandre Berard [Univ. Lille I]
Lilian Besson [ENS Cachan, from Oct 2016]
Daniele Calandriello [Inria]
Ronan Fruit [Inria]
Pratik Gajane [Orange Labs, granted by CIFRE]
Guillaume Gautier [Inria and CNRS]
Jean-Bastien Grill [granted by ENS Paris]
Frédéric Guillou [Inria]
Tomáš Kocák [Inria, until Nov 2016]
Julien Perolat [Univ. Lille I]
Florian Strub [Univ. Lille I]
Romain Warlop [55]

Post-Doctoral Fellow

James Ridgway [Inria, from Oct 2016]

Visiting Scientists

Maryam Aziz [Northeastern University, from May 2016 until Aug 2016]
Kamyar Azizzadenesheli [University of California at Irvine, from Aug 2016]
Cricia Zilda Felicio Paixao [University Uberlandia, Brasil, until Aug 2016]
Yao Ma [University of Tokyo, Japan, until Mar 2016]
Aristide Tossou [Chalmers University, Sweden]

Administrative Assistant

Amelie Supervielle [Inria]

Others

Mehdi Abbana Bennani [Univ. Lille III, intern, from Jun 2016 until Aug 2016]
Remi Bardenet [CNRS, Researcher]
Pierre Chainais [Ecole Centrale de Lille, Associate professor, HDR]
Pierre-Victor Chaumier [Inria, intern, from Feb 2016 until Jun 2016]
Quentin Coët [Univ. Lille I, intern, from Mar 2016 until Aug 2016]
Jean-Benoît Delbrouck [Univ. Lille I, intern, from Mar 2016 until Aug 2016]
Eddy El Khatib [Univ. Lille I, intern, from Apr 2016 until Jul 2016]

2. Overall Objectives

2.1. Presentation

SEQUEL means "Sequential Learning". As such, SEQUEL focuses on the task of learning in artificial systems (either hardware, or software) that gather information along time. Such systems are named *(learning) agents* (or learning machines) in the following. These data may be used to estimate some parameters of a model, which in turn, may be used for selecting actions in order to perform some long-term optimization task.

For the purpose of model building, the agent needs to represent information collected so far in some compact form and use it to process newly available data.

The acquired data may result from an observation process of an agent in interaction with its environment (the data thus represent a perception). This is the case when the agent makes decisions (in order to attain a certain objective) that impact the environment, and thus the observation process itself.

Hence, in SEQUEL, the term **sequential** refers to two aspects:

- The **sequential acquisition of data**, from which a model is learned (supervised and non supervised learning),
- the **sequential decision making task**, based on the learned model (reinforcement learning).

Examples of sequential learning problems include:

Supervised learning tasks deal with the prediction of some response given a certain set of observations of input variables and responses. New sample points keep on being observed.

Unsupervised learning tasks deal with clustering objects, these latter making a flow of objects. The (unknown) number of clusters typically evolves during time, as new objects are observed.

Reinforcement learning tasks deal with the control (a policy) of some system which has to be optimized (see [67]). We do not assume the availability of a model of the system to be controlled.

In all these cases, we mostly assume that the process can be considered stationary for at least a certain amount of time, and slowly evolving.

We wish to have any-time algorithms, that is, at any moment, a prediction may be required/an action may be selected making full use, and hopefully, the best use, of the experience already gathered by the learning agent.

The perception of the environment by the learning agent (using its sensors) is generally neither the best one to make a prediction, nor to take a decision (we deal with Partially Observable Markov Decision Problem). So, the perception has to be mapped in some way to a better, and relevant, state (or input) space.

Finally, an important issue of prediction regards its evaluation: how wrong may we be when we perform a prediction? For real systems to be controlled, this issue can not be simply left unanswered.

To sum-up, in SEQUEL, the main issues regard:

- the learning of a model: we focus on models that map some input space $\mathbb{R}^P$ to $\mathbb{R}$,
- the observation to state mapping,
- the choice of the action to perform (in the case of sequential decision problem),
- the performance guarantees,
- the implementation of usable algorithms,

all that being understood in a *sequential* framework.

3. Research Program

3.1. In Short

SEQUEL is primarily grounded on two domains:

- the problem of decision under uncertainty,
- statistical analysis and statistical learning, which provide the general concepts and tools to solve this problem.

To help the reader who is unfamiliar with these questions, we briefly present key ideas below.

3.2. Decision-making Under Uncertainty

The phrase "Decision under uncertainty" refers to the problem of taking decisions when we do not have a full knowledge neither of the situation, nor of the consequences of the decisions, as well as when the consequences of decision are non deterministic.

We introduce two specific sub-domains, namely the Markov decision processes which models sequential decision problems, and bandit problems.

3.2.1. Reinforcement Learning

Sequential decision processes occupy the heart of the SEQUEL project; a detailed presentation of this problem may be found in Puterman's book [65].

A Markov Decision Process (MDP) is defined as the tuple $(\mathcal{X}, \mathcal{A}, P, r)$ where $\mathcal{X}$ is the state space, $\mathcal{A}$ is the action space, P is the probabilistic transition kernel, and $r : \mathcal{X} \times \mathcal{A} \times \mathcal{X} \to I\!\!R$ is the reward function. For the sake of simplicity, we assume in this introduction that the state and action spaces are finite. If the current state (at time t) is $x \in \mathcal{X}$ and the chosen action is $a \in \mathcal{A}$, then the Markov assumption means that the transition probability to a new state $x' \in \mathcal{X}$ (at time $t+1$) only depends on (x, a). We write $p(x'|x, a)$ the corresponding transition probability. During a transition $(x, a) \to x'$, a reward $r(x, a, x')$ is incurred.

In the MDP $(\mathcal{X}, \mathcal{A}, P, r)$, each initial state x_0 and action sequence $a_0, a_1, ...$ gives rise to a sequence of states $x_1, x_2, ...$, satisfying $\mathbb{P}\left(x_{t+1} = x'|x_t = x, a_t = a\right) = p(x'|x, a)$, and rewards [0] $r_1, r_2, ...$ defined by $r_t = r(x_t, a_t, x_{t+1})$.

The history of the process up to time t is defined to be $H_t = (x_0, a_0, ..., x_{t-1}, a_{t-1}, x_t)$. A policy π is a sequence of functions $\pi_0, \pi_1, ...$, where π_t maps the space of possible histories at time t to the space of probability distributions over the space of actions $\mathcal{A}$. To follow a policy means that, in each time step, we assume that the process history up to time t is $x_0, a_0, ..., x_t$ and the probability of selecting an action a is equal to $\pi_t(x_0, a_0, ..., x_t)(a)$. A policy is called stationary (or Markovian) if π_t depends only on the last visited state. In other words, a policy $\pi = (\pi_0, \pi_1, ...)$ is called stationary if $\pi_t(x_0, a_0, ..., x_t) = \pi_0(x_t)$ holds for all $t \geq 0$. A policy is called deterministic if the probability distribution prescribed by the policy for any history is concentrated on a single action. Otherwise it is called a stochastic policy.

We move from an MD process to an MD problem by formulating the goal of the agent, that is what the sought policy π has to optimize? It is very often formulated as maximizing (or minimizing), in expectation, some functional of the sequence of future rewards. For example, an usual functional is the infinite-time horizon sum of discounted rewards. For a given (stationary) policy π, we define the value function $V^\pi(x)$ of that policy π at a state $x \in \mathcal{X}$ as the expected sum of discounted future rewards given that we state from the initial state x and follow the policy π:

$$V^\pi(x) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t | x_0 = x, \pi\right], \tag{9}$$

where $\mathbb{E}$ is the expectation operator and $\gamma \in (0, 1)$ is the discount factor. This value function V^π gives an evaluation of the performance of a given policy π. Other functionals of the sequence of future rewards may be considered, such as the undiscounted reward (see the stochastic shortest path problems [64]) and average reward settings. Note also that, here, we considered the problem of maximizing a reward functional, but a formulation in terms of minimizing some cost or risk functional would be equivalent.

[0] Note that for simplicity, we considered the case of a deterministic reward function, but in many applications, the reward r_t itself is a random variable.

In order to maximize a given functional in a sequential framework, one usually applies Dynamic Programming (DP) [62], which introduces the optimal value function $V^*(x)$, defined as the optimal expected sum of rewards when the agent starts from a state x. We have $V^*(x) = \sup_\pi V^\pi(x)$. Now, let us give two definitions about policies:

- We say that a policy π is optimal, if it attains the optimal values $V^*(x)$ for any state $x \in \mathcal{X}$, *i.e.*, if $V^\pi(x) = V^*(x)$ for all $x \in \mathcal{X}$. Under mild conditions, deterministic stationary optimal policies exist [63]. Such an optimal policy is written π^*.
- We say that a (deterministic stationary) policy π is greedy with respect to (w.r.t.) some function V (defined on $\mathcal{X}$) if, for all $x \in \mathcal{X}$,

$$\pi(x) \in \arg\max_{a\in\mathcal{A}} \sum_{x'\in\mathcal{X}} p(x'|x,a)\left[r(x,a,x') + \gamma V(x')\right].$$

  where $\arg\max_{a\in\mathcal{A}} f(a)$ is the set of $a \in \mathcal{A}$ that maximizes $f(a)$. For any function V, such a greedy policy always exists because $\mathcal{A}$ is finite.

The goal of Reinforcement Learning (RL), as well as that of dynamic programming, is to design an optimal policy (or a good approximation of it).

The well-known Dynamic Programming equation (also called the Bellman equation) provides a relation between the optimal value function at a state x and the optimal value function at the successors states x' when choosing an optimal action: for all $x \in \mathcal{X}$,

$$V^*(x) = \max_{a\in\mathcal{A}} \sum_{x'\in\mathcal{X}} p(x'|x,a)\left[r(x,a,x') + \gamma V^*(x')\right]. \tag{10}$$

The benefit of introducing this concept of optimal value function relies on the property that, from the optimal value function V^*, it is easy to derive an optimal behavior by choosing the actions according to a policy greedy w.r.t. V^*. Indeed, we have the property that a policy greedy w.r.t. the optimal value function is an optimal policy:

$$\pi^*(x) \in \arg\max_{a\in\mathcal{A}} \sum_{x'\in\mathcal{X}} p(x'|x,a)\left[r(x,a,x') + \gamma V^*(x')\right]. \tag{11}$$

In short, we would like to mention that most of the reinforcement learning methods developed so far are built on one (or both) of the two following approaches ([68]):

- Bellman's dynamic programming approach, based on the introduction of the value function. It consists in learning a "good" approximation of the optimal value function, and then using it to derive a greedy policy w.r.t. this approximation. The hope (well justified in several cases) is that the performance V^π of the policy π greedy w.r.t. an approximation V of V^* will be close to optimality. This approximation issue of the optimal value function is one of the major challenges inherent to the reinforcement learning problem. **Approximate dynamic programming** addresses the problem of estimating performance bounds (*e.g.* the loss in performance $||V^* - V^\pi||$ resulting from using a policy π-greedy w.r.t. some approximation V- instead of an optimal policy) in terms of the approximation error $||V^* - V||$ of the optimal value function V^* by V. Approximation theory and Statistical Learning theory provide us with bounds in terms of the number of sample data used to represent the functions, and the capacity and approximation power of the considered function spaces.

- Pontryagin's maximum principle approach, based on sensitivity analysis of the performance measure w.r.t. some control parameters. This approach, also called **direct policy search** in the Reinforcement Learning community aims at directly finding a good feedback control law in a parameterized policy space without trying to approximate the value function. The method consists in estimating the so-called **policy gradient**, *i.e.* the sensitivity of the performance measure (the value function) w.r.t. some parameters of the current policy. The idea being that an optimal control problem is replaced by a parametric optimization problem in the space of parameterized policies. As such, deriving a policy gradient estimate would lead to performing a stochastic gradient method in order to search for a local optimal parametric policy.

Finally, many extensions of the Markov decision processes exist, among which the Partially Observable MDPs (POMDPs) is the case where the current state does not contain all the necessary information required to decide for sure of the best action.

3.2.2. Multi-arm Bandit Theory

Bandit problems illustrate the fundamental difficulty of decision making in the face of uncertainty: A decision maker must choose between what seems to be the best choice ("exploit"), or to test ("explore") some alternative, hoping to discover a choice that beats the current best choice.

The classical example of a bandit problem is deciding what treatment to give each patient in a clinical trial when the effectiveness of the treatments are initially unknown and the patients arrive sequentially. These bandit problems became popular with the seminal paper [66], after which they have found applications in diverse fields, such as control, economics, statistics, or learning theory.

Formally, a K-armed bandit problem ($K \geq 2$) is specified by K real-valued distributions. In each time step a decision maker can select one of the distributions to obtain a sample from it. The samples obtained are considered as rewards. The distributions are initially unknown to the decision maker, whose goal is to maximize the sum of the rewards received, or equivalently, to minimize the regret which is defined as the loss compared to the total payoff that can be achieved given full knowledge of the problem, *i.e.*, when the arm giving the highest expected reward is pulled all the time.

The name "bandit" comes from imagining a gambler playing with K slot machines. The gambler can pull the arm of any of the machines, which produces a random payoff as a result: When arm k is pulled, the random payoff is drawn from the distribution associated to k. Since the payoff distributions are initially unknown, the gambler must use exploratory actions to learn the utility of the individual arms. However, exploration has to be carefully controlled since excessive exploration may lead to unnecessary losses. Hence, to play well, the gambler must carefully balance exploration and exploitation. Auer *et al.* [61] introduced the algorithm UCB (Upper Confidence Bounds) that follows what is now called the "optimism in the face of uncertainty principle". Their algorithm works by computing upper confidence bounds for all the arms and then choosing the arm with the highest such bound. They proved that the expected regret of their algorithm increases at most at a logarithmic rate with the number of trials, and that the algorithm achieves the smallest possible regret up to some sub-logarithmic factor (for the considered family of distributions).

3.3. Statistical analysis of time series

Many of the problems of machine learning can be seen as extensions of classical problems of mathematical statistics to their (extremely) non-parametric and model-free cases. Other machine learning problems are founded on such statistical problems. Statistical problems of sequential learning are mainly those that are concerned with the analysis of time series. These problems are as follows.

3.3.1. Prediction of Sequences of Structured and Unstructured Data

Given a series of observations $x_1, \cdots, x_n$ it is required to give forecasts concerning the distribution of the future observations $x_{n+1}, x_{n+2}, \cdots$; in the simplest case, that of the next outcome x_{n+1}. Then x_{n+1} is revealed and the process continues. Different goals can be formulated in this setting. One can either make some assumptions on the probability measure that generates the sequence $x_1, \cdots, x_n, \cdots$, such as that the outcomes are independent and identically distributed (i.i.d.), or that the sequence is a Markov chain, that it is a stationary process, etc. More generally, one can assume that the data is generated by a probability measure that belongs to a certain set $\mathcal{C}$. In these cases the goal is to have the discrepancy between the predicted and the "true" probabilities to go to zero, if possible, with guarantees on the speed of convergence.

Alternatively, rather than making some assumptions on the data, one can change the goal: the predicted probabilities should be asymptotically as good as those given by the best reference predictor from a certain pre-defined set.

Another dimension of complexity in this problem concerns the nature of observations x_i. In the simplest case, they come from a finite space, but already basic applications often require real-valued observations. Moreover, function or even graph-valued observations often arise in practice, in particular in applications concerning Web data. In these settings estimating even simple characteristics of probability distributions of the future outcomes becomes non-trivial, and new learning algorithms for solving these problems are in order.

3.3.2. Hypothesis testing

Given a series of observations of $x_1, \cdots, x_n, \cdots$ generated by some unknown probability measure μ, the problem is to test a certain given hypothesis H_0 about μ, versus a given alternative hypothesis H_1. There are many different examples of this problem. Perhaps the simplest one is testing a simple hypothesis "μ is Bernoulli i.i.d. measure with probability of 0 equals 1/2" versus "μ is Bernoulli i.i.d. with the parameter different from 1/2". More interesting cases include the problems of model verification: for example, testing that μ is a Markov chain, versus that it is a stationary ergodic process but not a Markov chain. In the case when we have not one but several series of observations, we may wish to test the hypothesis that they are independent, or that they are generated by the same distribution. Applications of these problems to a more general class of machine learning tasks include the problem of feature selection, the problem of testing that a certain behaviour (such as pulling a certain arm of a bandit, or using a certain policy) is better (in terms of achieving some goal, or collecting some rewards) than another behaviour, or than a class of other behaviours.

The problem of hypothesis testing can also be studied in its general formulations: given two (abstract) hypothesis H_0 and H_1 about the unknown measure that generates the data, find out whether it is possible to test H_0 against H_1 (with confidence), and if yes then how can one do it.

3.3.3. Change Point Analysis

A stochastic process is generating the data. At some point, the process distribution changes. In the "offline" situation, the statistician observes the resulting sequence of outcomes and has to estimate the point or the points at which the change(s) occurred. In online setting, the goal is to detect the change as quickly as possible.

These are the classical problems in mathematical statistics, and probably among the last remaining statistical problems not adequately addressed by machine learning methods. The reason for the latter is perhaps in that the problem is rather challenging. Thus, most methods available so far are parametric methods concerning piece-wise constant distributions, and the change in distribution is associated with the change in the mean. However, many applications, including DNA analysis, the analysis of (user) behaviour data, etc., fail to comply with this kind of assumptions. Thus, our goal here is to provide completely non-parametric methods allowing for any kind of changes in the time-series distribution.

3.3.4. Clustering Time Series, Online and Offline

The problem of clustering, while being a classical problem of mathematical statistics, belongs to the realm of unsupervised learning. For time series, this problem can be formulated as follows: given several samples $x^1 = (x_1^1, \cdots, x_{n_1}^1), \cdots, x^N = (x_N^1, \cdots, x_{n_N}^N)$, we wish to group similar objects together. While this is of course not a precise formulation, it can be made precise if we assume that the samples were generated by k different distributions.

The online version of the problem allows for the number of observed time series to grow with time, in general, in an arbitrary manner.

3.3.5. Online Semi-Supervised Learning

Semi-supervised learning (SSL) is a field of machine learning that studies learning from both labeled and unlabeled examples. This learning paradigm is extremely useful for solving real-world problems, where data is often abundant but the resources to label them are limited.

Furthermore, *online* SSL is suitable for adaptive machine learning systems. In the classification case, learning is viewed as a repeated game against a potentially adversarial nature. At each step t of this game, we observe an example $\mathbf{x_t}$, and then predict its label $\widehat{y}_t$.

The challenge of the game is that we only exceptionally observe the true label y_t. In the extreme case, which we also study, only a handful of labeled examples are provided in advance and set the initial bias of the system while unlabeled examples are gathered online and update the bias continuously. Thus, if we want to adapt to changes in the environment, we have to rely on indirect forms of feedback, such as the structure of data.

3.3.6. Online Kernel and Graph-Based Methods

Large-scale kernel ridge regression is limited by the need to store a large kernel matrix. Similarly, large-scale graph-based learning is limited by storing the graph Laplacian. Furthermore, if the data come online, at some point no finite storage is sufficient and per step operations become slow.

Our challenge is to design sparsification methods that give guaranteed approximate solutions with a reduced storage requirements.

4. Application Domains

4.1. Sequential decision making under uncertainty and prediction

The spectrum of applications of our research is very wide: it ranges from the core of our research, that is sequential decision making under uncertainty, to the application of components used to solve this decision making problem.

To be more specific, we work on computational advertizing and recommandation systems; these problems are considered as a sequential matching problem in which resources available in a limited amount have to be matched to meet some users' expectations. The sequential approach we advocate paves the way to better tackle the cold-start problem, and non stationary environments. More generally, these approaches are applied to the optimization of budgeted resources under uncertainty, in a time-varying environment, including constraints on computational times (typically, a decision has to be made in less than 1 ms in a recommandation system). An other field of applications of our research is related to education which we consider as a sequential matching problem between a student, and educational contents.

The algorithms to solve these tasks heavily rely on tools from machine learning, statistics, and optimization. Henceforth, we also apply our work to more classical supervised learning, and prediction tasks, as well as unsupervised learning tasks. The whole range of methods is used, from decision forests, to kernel methods, to deep learning. For instance, we have recently used deep learning on images. We also have a line of works related to software development studying how machine learning can improve the quality of software being developed. More generally, we apply our research to data science.

5. Highlights of the Year

5.1. Highlights of the Year

- Grill, Valko & Munos gave an oral presentation at NIPS. Oral presentations at NIPS are rare: out of 2500+ submissions, only 1.8% are presented orally.

- Using a deep learning approach (sparse denoising autoencoders), Strub, Mary & Gaudel have obtained the best ever published results on the data from the Netflix challenge on recommendation systems. 10 years ago, such an achievement was worth 1M$.

6. New Software and Platforms

6.1. BAC

Bayesian Policy Gradient and Actor-Critic Algorithms
KEYWORDS: Machine learning - Incremental learning - Policy Learning
FUNCTIONAL DESCRIPTION

To address this issue, we proceed to supplement our Bayesian policy gradient framework with a new actor-critic learning model in which a Bayesian class of non-parametric critics, based on Gaussian process temporal difference learning, is used. Such critics model the action-value function as a Gaussian process, allowing Bayes' rule to be used in computing the posterior distribution over action-value functions, conditioned on the observed data. Appropriate choices of the policy parameterization and of the prior covariance (kernel) between action-values allow us to obtain closed-form expressions for the posterior distribution of the gradient of the expected return with respect to the policy parameters. We perform detailed experimental comparisons of the proposed Bayesian policy gradient and actor-critic algorithms with classic Monte-Carlo based policy gradient methods, as well as with each other, on a number of reinforcement learning problems.

- Contact: Michal Valko
- URL: https://team.inria.fr/sequel/Software/BAC/

6.2. Collaborative Filtering Network

KEYWORDS: Recommender system - Neural networks - Deep learning
FUNCTIONAL DESCRIPTION

Recommendation systems advise users on which items (movies, musics, books etc.) they are more likely to be interested in. A good recommendation system may dramatically increase the amount of sales of a firm or retain customers. For instance, 80% of movies watched on Netflix come from the recommender system of the company. Colaborative Filtering (CF) aims at recommending an item to a user by predicting how a user would rate this item. To do so, the feedback of one user on some items is combined with the feedback of all other users on all items to predict a new rating. For instance, if someone rated a few books, CF objective is to estimate the ratings he would have given to thousands of other books by using the ratings of all the other readers.

The following module tackles Collaborative Filtering tasks by using a novel approach based on neural networks (sparse denoising autoencoders). In a few words, the module lets the user train neural networks to predict unknown entries in a history files.

The input files are classic csv files. The output files can either be the full matrix of ratings and/or the network weights. The root mean square error is computed to assess the quality of the training.

This module is based on Lua/Torch Framework. It works on both CPU/GPU and it is multithreaded.

- Contact: Florian Strub
- URL: https://github.com/fstrub95/Autoencoders_cf

7. New Results

7.1. Decision-making Under Uncertainty

7.1.1. Reinforcement Learning

Analysis of Classification-based Policy Iteration Algorithms, [20] We introduce a variant of the classification-based approach to policy iteration which uses a cost-sensitive loss function weighting each classification mistake by its actual regret, that is, the difference between the action-value of the greedy action and of the action chosen by the classifier. For this algorithm, we provide a full finite-sample analysis. Our results state a performance bound in terms of the number of policy improvement steps, the number of rollouts used in each iteration, the capacity of the considered policy space (classifier), and a capacity measure which indicates how well the policy space can approximate policies that are greedy with respect to any of its members. The analysis reveals a tradeoff between the estimation and approximation errors in this classification-based policy iteration setting. Furthermore it confirms the intuition that classification-based policy iteration algorithms could be favorably compared to value-based approaches when the policies can be approximated more easily than their corresponding value functions. We also study the consistency of the algorithm when there exists a sequence of policy spaces with increasing capacity.

Reinforcement Learning of POMDPs using Spectral Methods, [23]

We propose a new reinforcement learning algorithm for partially observable Markov decision processes (POMDP) based on spectral decomposition methods. While spectral methods have been previously employed for consistent learning of (passive) latent variable models such as hidden Markov models, POMDPs are more challenging since the learner interacts with the environment and possibly changes the future observations in the process. We devise a learning algorithm running through episodes, in each episode we employ spectral techniques to learn the POMDP parameters from a trajectory generated by a fixed policy. At the end of the episode, an optimization oracle returns the optimal memoryless planning policy which maximizes the expected reward based on the estimated POMDP model. We prove an order-optimal regret bound w.r.t. the optimal memoryless policy and efficient scaling with respect to the dimensionality of observation and action spaces.

Bayesian Policy Gradient and Actor-Critic Algorithms, [15]

Policy gradient methods are reinforcement learning algorithms that adapt a parameterized policy by following a performance gradient estimate. Many conventional policy gradient methods use Monte-Carlo techniques to estimate this gradient. The policy is improved by adjusting the parameters in the direction of the gradient estimate. Since Monte-Carlo methods tend to have high variance, a large number of samples is required to attain accurate estimates, resulting in slow convergence. In this paper, we first propose a Bayesian framework for policy gradient, based on modeling the policy gradient as a Gaussian process. This reduces the number of samples needed to obtain accurate gradient estimates. Moreover, estimates of the natural gradient as well as a measure of the uncertainty in the gradient estimates, namely, the gradient covariance, are provided at little extra cost. Since the proposed Bayesian framework considers system trajectories as its basic observable unit, it does not require the dynamics within trajectories to be of any particular form, and thus, can be easily extended to partially observable problems. On the downside, it cannot take advantage of the Markov property when the system is Markovian. To address this issue, we proceed to supplement our Bayesian policy gradient framework with a new actor-critic learning model in which a Bayesian class of non-parametric critics, based on Gaussian process temporal difference learning, is used. Such critics model the action-value function as a Gaussian process, allowing Bayes' rule to be used in computing the posterior distribution over action-value functions, conditioned on the observed data. Appropriate choices of the policy parameterization and of the prior covariance (kernel) between action-values allow us to obtain closed-form expressions for the posterior distribution of the gradient of the expected return with respect to the policy parameters. We perform detailed experimental comparisons of the proposed Bayesian policy gradient and actor-critic algorithms with classic Monte-Carlo based policy gradient methods, as well as with each other, on a number of reinforcement learning problems.

7.1.2. Multi-arm Bandit Theory

Improved Learning Complexity in Combinatorial Pure Exploration Bandits, [32]

We study the problem of combinatorial pure exploration in the stochastic multi-armed bandit problem. We first construct a new measure of complexity that provably characterizes the learning performance of the algorithms we propose for the fixed confidence and the fixed budget setting. We show that this complexity is never higher than the one in existing work and illustrate a number of configurations in which it can be significantly smaller.

While in general this improvement comes at the cost of increased computational complexity, we provide a series of examples , including a planning problem, where this extra cost is not significant.

Online learning with noisy side observations, [43]

We propose a new partial-observability model for online learning problems where the learner, besides its own loss, also observes some noisy feedback about the other actions, depending on the underlying structure of the problem. We represent this structure by a weighted directed graph, where the edge weights are related to the quality of the feedback shared by the connected nodes. Our main contribution is an efficient algorithm that guarantees a regret of $O(\sqrt{\alpha * T})$ after T rounds, where α * is a novel graph property that we call the effective independence number. Our algorithm is completely parameter-free and does not require knowledge (or even estimation) of α *. For the special case of binary edge weights, our setting reduces to the partial-observability models of Mannor & Shamir (2011) and Alon et al. (2013) and our algorithm recovers the near-optimal regret bounds.

Online learning with Erdös-Rényi side-observation graphs, [42]

We consider adversarial multi-armed bandit problems where the learner is allowed to observe losses of a number of arms beside the arm that it actually chose. We study the case where all non-chosen arms reveal their loss with an unknown probability rt, independently of each other and the action of the learner. Moreover, we allow rt to change in every round t, which rules out the possibility of estimating rt by a well-concentrated sample average. We propose an algorithm which operates under the assumption that rt is large enough to warrant at least one side observation with high probability. We show that after T rounds in a bandit problem with N arms, the expected regret of our algorithm is of order O(sqrt(sum(t=1)T (1/rt) log N)), given that rt less than log T / (2N-2) for all t. All our bounds are within logarithmic factors of the best achievable performance of any algorithm that is even allowed to know exact values of rt.

Revealing graph bandits for maximizing local influence, [27]

We study a graph bandit setting where the objective of the learner is to detect the most influential node of a graph by requesting as little information from the graph as possible. One of the relevant applications for this setting is marketing in social networks, where the marketer aims at finding and taking advantage of the most influential customers. The existing approaches for bandit problems on graphs require either partial or complete knowledge of the graph. In this paper, we do not assume any knowledge of the graph, but we consider a setting where it can be gradually discovered in a sequential and active way. At each round, the learner chooses a node of the graph and the only information it receives is a stochastic set of the nodes that the chosen node is currently influencing. To address this setting, we propose BARE, a bandit strategy for which we prove a regret guarantee that scales with the detectable dimension, a problem dependent quantity that is often much smaller than the number of nodes.

Algorithms for Differentially Private Multi-Armed Bandits, [50]

We present differentially private algorithms for the stochastic Multi-Armed Bandit (MAB) problem. This is a problem for applications such as adaptive clinical trials, experiment design, and user-targeted advertising where private information is connected to individual rewards. Our major contribution is to show that there exist (ϵ, δ) differentially private variants of Upper Confidence Bound algorithms which have optimal regret, $O(\epsilon^{-1} + \log T)$. This is a significant improvement over previous results, which only achieve poly-log regret $O(\epsilon^{-2} \log^2 T)$, because of our use of a novel interval-based mechanism. We also substantially improve the bounds of previous family of algorithms which use a continual release mechanism. Experiments clearly validate our theoretical bounds.

On the Complexity of Best Arm Identification in Multi-Armed Bandit Models, [17]

The stochastic multi-armed bandit model is a simple abstraction that has proven useful in many different contexts in statistics and machine learning. Whereas the achievable limit in terms of regret minimization is now well known, our aim is to contribute to a better understanding of the performance in terms of identifying the m best arms. We introduce generic notions of complexity for the two dominant frameworks considered in the literature: fixed-budget and fixed-confidence settings. In the fixed-confidence setting, we provide the first known distribution-dependent lower bound on the complexity that involves information-theoretic quantities and holds when m is larger than 1 under general assumptions. In the specific case of two armed-bandits, we derive refined lower bounds in both the fixed-confidence and fixed-budget settings, along with matching algorithms for Gaussian and Bernoulli bandit models. These results show in particular that the complexity of the fixed-budget setting may be smaller than the complexity of the fixed-confidence setting, contradicting the familiar behavior observed when testing fully specified alternatives. In addition, we also provide improved sequential stopping rules that have guaranteed error probabilities and shorter average running times. The proofs rely on two technical results that are of independent interest : a deviation lemma for self-normalized sums (Lemma 19) and a novel change of measure inequality for bandit models (Lemma 1).

Optimal Best Arm Identification with Fixed Confidence, [33]

We give a complete characterization of the complexity of best-arm identification in one-parameter bandit problems. We prove a new, tight lower bound on the sample complexity. We propose the ‘Track-and-Stop’ strategy, which we prove to be asymptotically optimal. It consists in a new sampling rule (which tracks the optimal proportions of arm draws highlighted by the lower bound) and in a stopping rule named after Chernoff, for which we give a new analysis.

On Explore-Then-Commit Strategies, [35]

We study the problem of minimising regret in two-armed bandit problems with Gaussian rewards. Our objective is to use this simple setting to illustrate that strategies based on an exploration phase (up to a stopping time) followed by exploitation are necessarily suboptimal. The results hold regardless of whether or not the difference in means between the two arms is known. Besides the main message, we also refine existing deviation inequalities, which allow us to design fully sequential strategies with finite-time regret guarantees that are (a) asymptotically optimal as the horizon grows and (b) order-optimal in the minimax sense. Furthermore we provide empirical evidence that the theory also holds in practice and discuss extensions to non-gaussian and multiple-armed case.

7.1.3. Recommendation systems

Scalable explore-exploit Collaborative Filtering, [39]

Recommender Systems (RS) aim at suggesting to users one or several items in which they might have interest. These systems have to update themselves as users provide new ratings, but also as new users/items enter the system. While this adaptation makes recommendation an intrinsically sequential task, most researches about RS based on Collaborative Filtering are omitting this fact, as well as the ensuing exploration/exploitation dilemma: should the system recommend items which bring more information about the users (explore), or should it try to get an immediate feedback as high as possible (exploit)? Recently, a few approaches were proposed to solve that dilemma, but they do not meet requirements to scale up to real life applications which is a crucial point as the number of items available on RS and the number of users in these systems explode. In this paper, we present an explore-exploit Collaborative Filtering RS which is both efficient and scales well. Extensive experiments on some of the largest available real-world datasets show that the proposed approach performs accurate personalized recommendations in less than a millisecond per recommendation, which makes it a good candidate for true applications.

Large-scale Bandit Recommender System, [38]

The main target of Recommender Systems (RS) is to propose to users one or several items in which they might be interested. However, as users provide more feedback, the recommendation process has to take these new data into consideration. The necessity of this update phase makes recommendation an intrinsically sequential task. A few approaches were recently proposed to address this issue, but they do not meet the need to scale up to real life applications. In this paper , we present a Collaborative Filtering RS method based on Matrix Factorization and Multi-Armed Bandits. This approach aims at good recommendations with a narrow computation time. Several experiments on large datasets show that the proposed approach performs personalized recommendations in less than a millisecond per recommendation.

Sequential Collaborative Ranking Using (No-)Click Implicit Feedback, [40] We study Recommender Systems in the context where they suggest a list of items to users. Several crucial issues are raised in such a setting: first, identify the relevant items to recommend; second, account for the feedback given by the user after he clicked and rated an item; third, since new feedback arrive into the system at any moment, incorporate such information to improve future recommendations. In this paper, we take these three aspects into consideration and present an approach handling click/no-click feedback information. Experiments on real-world datasets show that our approach outperforms state of the art algorithms.

Hybrid Recommender System based on Autoencoders, [49]

A standard model for Recommender Systems is the Matrix Completion setting: given partially known matrix of ratings given by users (rows) to items (columns), infer the unknown ratings. In the last decades, few attempts where done to handle that objective with Neural Networks, but recently an architecture based on Autoencoders proved to be a promising approach. In current paper, we enhanced that architecture (i) by using a loss function adapted to input data with missing values, and (ii) by incorporating side information. The experiments demonstrate that while side information only slightly improve the test error averaged on all users/items, it has more impact on cold users/items.

Compromis exploration-exploitation pour système de recommandation à grande échelle, [53]

Les systèmes de recommandation recommandent à des utilisateurs un ou des produits qui pourraient les intéresser. La recommandation se fonde sur les retours des utilisateurs par le passé, lors des précédentes recommandations. La recommandation est donc un problème séquentiel et le système de recommandation recommande (i) pour obtenir une bonne récompense, mais aussi (ii) pour mieux cerné l'utilisateur/les produits et ainsi obtenir de meilleures récompenses par la suite. Quelques approches récentes ciblent ce double objectif mais elles sont trop gourmandes en temps de calcul pour s'appliquer à certaines applications de la vie réelle. Dans cet article, nous présentons un système de recommandation fondé sur la factorisation de matrice et les bandits manchots. Plusieurs expériences sur de grandes base de données montrent que l'approche proposée fournit de bonnes recommendations en moins d'une milli-seconde par recommandation.

Filtrage Collaboratif Hybride avec des Auto-encodeurs, [54]

Le filtrage collaboratif (CF) exploite les retours des utilisateurs pour leur fournir des recommandations personnalisées. Lorsque ces algorithmes ont accès à des informations complémentaires, ils ont de meilleurs résultats et gèrent plus efficacement le démarrage à froid. Bien que les réseaux de neurones (NN) remportent de nombreux succès en traitement d'images, ils ont reçu beaucoup moins d'attention dans la communauté du CF. C'est d'autant plus surprenant que les NN apprennent comme les algorithme de CF une représentation latente des données. Dans cet article, nous introduisons une architecture de NN adaptée au CF (nommée CFN) qui prend en compte la parcimonie des données et les informations complémentaires. Nous montrons empiriquement sur les bases de données MovieLens et Douban que CFN bât l'état de l'art et profite des informations complémentaires. Nous fournissons une implémentation de l'algorithme sous forme d'un plugin pour Torch.

7.1.4. Nonparametric statistics of time series

Things Bayes can't do, [48]

The problem of forecasting conditional probabilities of the next event given the past is consideredin a general probabilistic setting. Given an arbitrary (large, uncountable) set C of predictors, we would like to construct a single predictor that performs asymptotically as well as the best predictor in C, on any data. Here we show that there are sets C for which such predictors exist, but none of them is a Bayesian predictor with a prior concentrated on C.In other words, there is a predictor with sublinear regret, but every Bayesian predictor must have a linear regret. This negative finding is in sharp contrast with previous resultsthat establish the opposite for the case when one of the predictors in C achieves asymptotically vanishing error.In such a case, if there is a predictor that achieves asymptotically vanishing error for any measure in C, then there is a Bayesian predictor that also has this property, and whose prior is concentrated on (a countable subset of) C.

7.1.5. Imitation and Inverse Reinforcement Learning

Score-based Inverse Reinforcement Learning, [29] This paper reports theoretical and empirical results obtained for the score-based Inverse Reinforcement Learning (IRL) algorithm. It relies on a non-standard setting for IRL consisting of learning a reward from a set of globally scored trajec-tories. This allows using any type of policy (optimal or not) to generate trajectories without prior knowledge during data collection. This way, any existing database (like logs of systems in use) can be scored a posteriori by an expert and used to learn a reward function. Thanks to this reward function, it is shown that a near-optimal policy can be computed. Being related to least-square regression, the algorithm (called SBIRL) comes with theoretical guarantees that are proven in this paper. SBIRL is compared to standard IRL algorithms on synthetic data showing that annotations do help under conditions on the quality of the trajectories. It is also shown to be suitable for real-world applications such as the optimisation of a spoken dialogue system.

7.1.6. Stochastic Games

Blazing the trails before beating the path: Sample-efficient Monte-Carlo planning, [37]

You are a robot and you live in a Markov decision process (MDP) with a finite or an infinite number of transitions from state-action to next states. You got brains and so you plan before you act. Luckily, your roboparents equipped you with a generative model to do some Monte-Carlo planning. The world is waiting for you and you have no time to waste. You want your planning to be efficient. Sample-efficient. Indeed, you want to exploit the possible structure of the MDP by exploring only a subset of states reachable by following near-optimal policies. You want guarantees on sample complexity that depend on a measure of the quantity of near-optimal states. You want something, that is an extension of Monte-Carlo sampling (for estimating an expectation) to problems that alternate maximization (over actions) and expectation (over next states). But you do not want to StOP with exponential running time, you want something simple to implement and computationally efficient. You want it all and you want it now. You want TrailBlazer.

Maximin Action Identification: A New Bandit Framework for Games, [34]

We study an original problem of pure exploration in a strategic bandit model motivated by Monte Carlo Tree Search. It consists in identifying the best action in a game, when the player may sample random outcomes of sequentially chosen pairs of actions. We propose two strategies for the fixed-confidence setting: Maximin-LUCB, based on lower-and upper-confidence bounds; and Maximin-Racing, which operates by successively eliminating the sub-optimal actions. We discuss the sample complexity of both methods and compare their performance empirically. We sketch a lower bound analysis, and possible connections to an optimal algorithm.

7.2. Statistical analysis of time series

7.2.1. Change Point Analysis

Nonparametric multiple change point estimation in highly dependent time series, [18]

Given a heterogeneous time-series sample, the objective is to find points in time, called change points, where the probability distribution generating the data has changed. The data are assumed to have been generated by arbitrary unknown stationary ergodic distributions. No modelling, independence or mixing assumptions are made. A novel, computationally efficient, nonparametric method is proposed, and is shown to be asymptotically consistent in this general framework. The theoretical results are complemented with experimental evaluations.

7.2.2. Clustering Time Series, Online and Offline

Consistent Algorithms for Clustering Time Series, [19]

The problem of clustering is considered for the case where every point is a time series. The time series are either given in one batch (offline setting), or they are allowed to grow with time and new time series can be added along the way (online setting). We propose a natural notion of consistency for this problem, and show that there are simple, com-putationally efficient algorithms that are asymptotically consistent under extremely weak assumptions on the distributions that generate the data. The notion of consistency is as follows. A clustering algorithm is called consistent if it places two time series into the same cluster if and only if the distribution that generates them is the same. In the considered framework the time series are allowed to be highly dependent, and the dependence can have arbitrary form. If the number of clusters is known, the only assumption we make is that the (marginal) distribution of each time series is stationary ergodic. No parametric, memory or mixing assumptions are made. When the number of clusters is unknown, stronger assumptions are provably necessary, but it is still possible to devise nonparametric algorithms that are consistent under very general conditions. The theoretical findings of this work are illustrated with experiments on both synthetic and real data.

7.2.3. Automata Learning

PAC learning of Probabilistic Automaton based on the Method of Moments, [36]

Probabilitic Finite Automata (PFA) are gener-ative graphical models that define distributions with latent variables over finite sequences of symbols, a.k.a. stochastic languages. Traditionally , unsupervised learning of PFA is performed through algorithms that iteratively improves the likelihood like the Expectation-Maximization (EM) algorithm. Recently, learning algorithms based on the so-called Method of Moments (MoM) have been proposed as a much faster alternative that comes with PAC-style guarantees. However, these algorithms do not ensure the learnt automata to model a proper distribution , limiting their applicability and preventing them to serve as an initialization to iterative algorithms. In this paper, we propose a new MoM-based algorithm with PAC-style guarantees that learns automata defining proper distributions. We assess its performances on synthetic problems from the PAutomaC challenge and real datasets extracted from Wikipedia against previous MoM-based algorithms and EM algorithm.

7.2.4. Online Kernel and Graph-Based Methods

Analysis of Nyström method with sequential ridge leverage score sampling, [26]

Large-scale kernel ridge regression (KRR) is limited by the need to store a large kernel matrix Kt. To avoid storing the entire matrix Kt, Nystro¨m methods subsample a subset of columns of the kernel matrix, and efficiently find an approximate KRR solution on the reconstructed Kt . The chosen subsampling distribution in turn affects the statistical and computational tradeoffs. For KRR problems, [15, 1] show that a sampling distribution proportional to the ridge leverage scores (RLSs) provides strong reconstruction guarantees for Kt. While exact RLSs are as difficult to compute as a KRR solution, we may be able to approximate them well enough. In this paper, we study KRR problems in a sequential setting and introduce the INK-ESTIMATE algorithm, that incrementally computes the RLSs estimates. INK-ESTIMATE maintains a small sketch of Kt, that at each step is used to compute an intermediate estimate of the RLSs. First, our sketch update does not require access to previously seen columns, and therefore a single pass over the kernel matrix is sufficient. Second, the algorithm requires a fixed, small space budget to run dependent only on the effective dimension of the kernel matrix. Finally, our sketch provides strong approximation guarantees on the distance $||Kt - Kt||^2$, and on the statistical risk of the approximate KRR solution at any time, because all our guarantees hold at any intermediate step.

7.3. Statistical Learning and Bayesian Analysis

7.3.1. Non-parametric methods for Function Approximation

Pliable rejection sampling, [30]

Rejection sampling is a technique for sampling from difficult distributions. However, its use is limited due to a high rejection rate. Common adaptive rejection sampling methods either work only for very specific distributions or without performance guarantees. In this paper, we present pliable rejection sampling (PRS), a new approach to rejection sampling, where we learn the sampling proposal using a kernel estimator. Since our method builds on rejection sampling, the samples obtained are with high probability i.i.d. and distributed according to f. Moreover, PRS comes with a guarantee on the number of accepted samples.

7.3.2. Non-parametric methods for functional supervised learning

Operator-valued Kernels for Learning from Functional Response Data, [16] In this paper we consider the problems of supervised classification and regression in the case where attributes and labels are functions: a data is represented by a set of functions, and the label is also a function. We focus on the use of reproducing kernel Hilbert space theory to learn from such functional data. Basic concepts and properties of kernel-based learning are extended to include the estimation of function-valued functions. In this setting, the representer theorem is restated, a set of rigorously defined infinite-dimensional operator-valued kernels that can be valuably applied when the data are functions is described, and a learning algorithm for nonlinear functional data analysis is introduced. The methodology is illustrated through speech and audio signal processing experiments.

7.3.3. Differential privacy

On the Differential Privacy of Bayesian Inference, [51]

We study how to communicate findings of Bayesian inference to third parties, while preserving the strong guarantee of differential privacy. Our main contributions are four different algorithms for private Bayesian inference on proba-bilistic graphical models. These include two mechanisms for adding noise to the Bayesian updates, either directly to the posterior parameters, or to their Fourier transform so as to preserve update consistency. We also utilise a recently introduced posterior sampling mechanism, for which we prove bounds for the specific but general case of discrete Bayesian networks; and we introduce a maximum-a-posteriori private mechanism. Our analysis includes utility and privacy bounds, with a novel focus on the influence of graph structure on privacy. Worked examples and experiments with Bayesian naïve Bayes and Bayesian linear regression illustrate the application of our mechanisms.

Algorithms for Differentially Private Multi-Armed Bandits, [50]

We present differentially private algorithms for the stochastic Multi-Armed Bandit (MAB) problem. This is a problem for applications such as adaptive clinical trials, experiment design, and user-targeted advertising where private information is connected to individual rewards. Our major contribution is to show that there exist (ϵ, δ) differentially private variants of Upper Confidence Bound algorithms which have optimal regret, $O(\epsilon^{-1} + \log T)$. This is a significant improvement over previous results, which only achieve poly-log regret $O(\epsilon^{-2} \log^2 T)$, because of our use of a novel interval-based mechanism. We also substantially improve the bounds of previous family of algorithms which use a continual release mechanism. Experiments clearly validate our theoretical bounds.

7.4. Applications

7.4.1. Spoken Dialogue Systems

Compact and Interpretable Dialogue State Representation with Genetic Sparse Distributed Memory, [28]

t User satisfaction is often considered as the objective that should be achieved by spoken dialogue systems. This is why, the reward function of Spoken Dialogue Systems (SDS) trained by Reinforcement Learning (RL) is often designed to reflect user satisfaction. To do so, the state space representation should be based on features capturing user satisfaction characteristics such as the mean speech recognition confidence score for instance. On the other hand, for deployment in industrial systems, there is a need for state representations that are understandable by system engineers. In this paper, we propose to represent the state space using a Genetic Sparse Distributed Memory. This is a state aggregation method computing state prototypes which are selected so as to lead to the best linear representation of the value function in RL. To do so, previous work on Genetic Sparse Distributed Memory for classification is adapted to the Reinforcement Learning task and a new way of building the prototypes is proposed. The approach is tested on a corpus of dialogues collected with an appointment scheduling system. The results are compared to a grid-based linear parametrisation. It is shown that learning is accelerated and made more memory efficient. It is also shown that the framework is calable in that it is possible to include many dialogue features in the representation, interpret the resulting policy and identify the most important dialogue features.

A Stochastic Model for Computer-Aided Human-Human Dialogue, [24] In this paper we introduce a novel model for computer-aided human-human dialogue. In this context, the computer aims at improving the outcome of a human-human task-oriented dialogue by intervening during the course of the interaction. While dialogue state and topic tracking in human-human dialogue have already been studied, few work has been devoted to the sequential part of the problem, where the impact of the system's actions on the future of the conversation is taken into account. This paper addresses this issue by first modelling human-human dialogue as a Markov Reward Process. The task of purposely taking part into the conversation is then optimised within the Linearly Solvable Markov Decision Process framework. Utterances of the Conversational Agent are seen as perturbations in this process, which aim at satisfying the user's long-term goals while keeping the conversation natural. Finally, results obtained by simulation suggest that such an approach is suitable for computer-aided human-human dialogue and is a first step towards three-party dialogue.

Learning Dialogue Dynamics with the Method of Moments, [25]

In this paper, we introduce a novel framework to encode the dynamics of dialogues into a probabilistic graphical model. Traditionally, Hidden Markov Models (HMMs) would be used to address this problem, involving a first step of hand-crafting to build a dialogue model (e.g. defining potential hidden states) followed by applying expectation-maximisation (EM) algorithms to refine it. Recently, an alternative class of algorithms based on the Method of Moments (MoM) has proven successful in avoiding issues of the EM-like algorithms such as convergence towards local optima, tractability issues, initialization issues or the lack of theoretical guarantees. In this work, we show that dialogues may be modeled by SP-RFA, a class of graphical models efficiently learnable within the MoM and directly usable in planning algorithms (such as reinforcement learning). Experiments are led on the Ubuntu corpus and dialogues are considered as sequences of dialogue acts, represented by their Latent Dirichlet Allocation (LDA) and Latent Semantic Analysis (LSA). We show that a MoM-based algorithm can learn a compact model of sequences of such acts.

7.4.2. Software development

Mutation-Based Graph Inference for Fault Localization, [45]

We present a new fault localization algorithm, called Vautrin, built on an approximation of causality based on call graphs. The approximation of causality is done using software mutants. The key idea is that if a mutant is killed by a test, certain call graph edges within a path between the mutation point and the failing test are likely causal. We evaluate our approach on the fault localization benchmark by Steimann et al. totaling 5,836 faults. The causal graphs are extracted from 88,732 nodes connected by 119,531 edges. Vautrin improves the fault localization effectiveness for all subjects of the benchmark. Considering the wasted effort at the method level, a classical fault localization evaluation metric, the improvement ranges from 3

A Large-scale Study of Call Graph-based Impact Prediction using Mutation Testing, [21]

In software engineering, impact analysis consists in predicting the software elements (e.g. modules, classes, methods) potentially impacted by a change in the source code. Impact analysis is required to optimize the testing effort. In this paper, we propose a framework to predict error propagation. Based on 10 open-source Java projects and 5 classical mutation operators, we create 17000 mutants and study how the error they introduce propagates. This framework enables us to analyze impact prediction based on four types of call graph. Our results show that the sophistication indeed increases completeness of impact prediction. However, and surprisingly to us, the most basic call graph gives the highest trade-off between precision and recall for impact prediction.

A Learning Algorithm for Change Impact Prediction, [44]

Change impact analysis (CIA) consists in predicting the impact of a code change in a software application. In this paper, the artifacts that are considered for CIA are methods of object-oriented software; the change under study is a change in the code of the method, the impact is the test methods that fail because of the change that has been performed. We propose LCIP, a learning algorithm that learns from past impacts to predict future impacts. To evaluate LCIP, we consider Java software applications that are strongly tested. We simulate 6000 changes and their actual impact through code mutations, as done in mutation testing. We find that LCIP can predict the impact with a precision of 74

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- contract with “500px”; PI: Romaric Gaudel.

  Title: Recommender System for Photos

  Duration: May 2016 – Oct. 2016 (6 months)

  Abstract: Recommender Systems aim at recommending items to users. Advances in that field are targeting more and more personalized recommendation. From a recommendation based on market segment to a recommendation based on individual user taste. From a recommendation based on user’s information to a recommendation based on any feedback from any user. From a recommendation based on logged data to a recommendation including latest trends... 500px is a Canadian company which is part of this trend. 500px offers solutions to store pictures online, to share pictures, and to browse among pictures exhibited by other users. Given the huge amount of pictures stored by 500px, users need help to find pictures which corresponds to their tastes. 500px offers several tools to filter the content presented to users. But the tools allowing exploration of the pictures landscape are not personalized, the selection is mostly based on the popularity of pictures/galleries. The most personalized recommendations are obtained by following other users: you see recent pictures of that users. But such recommendations requires you (i) to discover by yourself relevant users, (ii) to explicitly tag these users. The aim of the project is to scan state of the art in Collaborative Filtering and to design a tool which recommends pictures to users based on their implicit actions: given the list of followed users, famed pictures, commented pictures, browsed pictures, ..., infer user’s tastes and recommend to that user pictures and/or other user to look at. The system would also make use of informations on the pictures and of user profiles.

- contract with “Orange Labs”; PI: Philippe Preux

  Title: Sequential Learning and Decision Making under Partial Monitoring

  Duration: Oct. 2014 – Sep. 2017

  Abstract: In applications such as recommendation systems, or computational advertising, the return collected from the user is partial: (s)he clicks on one item, or no item at all. We study this setting in which only a “partial” information is gathered in particular how to learn to behave optimaly in such a setting.

- contract with “55”; PI: Jérémie Mary

  Title: Novel Learning and Exploration-Exploitation Methods for Effective Recommender Systems

  Duration: Oct. 2015 – Sep. 2018

  Abstract: In this Ph.D. thesis we intend to deal with this problem by developing novel and more sophisticated recommendation strategies in which the collection of data and the improvement of the performance are considered as a unique process, where the trade-off between the quality of the data and the performance of the recommendation strategy is optimized over time. This work also consider tensor methods (one layer of the tensor can be the time) with the goal to scale them at RS level.

- contract with “What a nice place” ; PI: Jérémie Mary

  Title: Deduplication of pictures

  Duration: Mar. 2016 – Jan. 2017

  Abstract: “What is nice place” is a start up which aggregates products from different sources in order to provide some home staging advises. Uniqueness of presence for the items in their database can be hard to achieve because of the differences over names and variations of a product. Here we build a classification and deduplication system based on deep neural networks. In this contract we received support from Inria Tech and transferred them some knowledge about deep neural networks.

- contract with “What a nice place” and “Leroy Merlin”; PI: Jérémie Mary

  Title: New Shopping Experience - Virtual Coach

  Duration: Jun. 2016 – Fev. 2017

  Abstract: The goal of this project is to use pictures of house interiors in order to propose automatically some products which would fit in nicely. The relations are learnt automatically using deep neural networks and recommendation systems techniques. We made a first version which focuses on lamps which is available for demonstration at https://whataniceplace.leroymerlin.fr/

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR BoB

Participant: Michal Valko.

- *Title*: Bayesian statistics for expensive models and tall data
- *Type*: National Research Agency
- *Coordinator*: CNRS (R. Bardenet)
- *Duration*: 2016-2020
- *Abstract*:

  Bayesian methods are a popular class of statistical algorithms for updating scientific beliefs. They turn data into decisions and models, taking into account uncertainty about models and their parameters. This makes Bayesian methods popular among applied scientists such as biologists, physicists, or engineers. However, at the heart of Bayesian analysis lie 1) repeated sweeps over the full dataset considered, and 2) repeated evaluations of the model that describes the observed physical process. The current trends to large-scale data collection and complex models thus raises two main issues. Experiments, observations, and numerical simulations in many areas of science nowadays generate terabytes of data, as does the LHC in particle physics for instance. Simultaneously, knowledge creation is becoming more and more data-driven, which requires new paradigms addressing how data are captured, processed, discovered, exchanged, distributed, and analyzed. For statistical algorithms to scale up, reaching a given performance must require as few iterations and as little access to data as possible. It is not only experimental measurements that are growing at a rapid pace. Cell biologists tend to have scarce data but large-scale models of tens of nonlinear differential equations to describe complex dynamics. In such settings, evaluating the model once requires numerically solving a large system of differential equations, which may take minutes for some tens of differential equations on today’s hardware. Iterative statistical processing that requires a million sequential runs of the model is thus out of the question. In this project, we tackle the fundamental cost-accuracy trade-off for Bayesian methods, in order to produce generic inference algorithms that scale favourably with the number of measurements in an experiment and the number of runs of a statistical model. We propose a collection of objectives with different risk-reward trade-offs to tackle these two goals. In particular, for experiments with large numbers of measurements, we further develop existing subsampling-based Monte Carlo methods, while developing a novel decision theory framework that includes data constraints. For expensive models, we build an ambitious programme around Monte Carlo methods that leverage determinantal processes, a rich class of probabilistic tools that lead to accurate inference with limited model evaluations. In short, using innovative techniques such as subsampling-based Monte Carlo and determinantal point processes, we propose in this project to push the boundaries of the applicability of Bayesian inference.

9.1.2. ANR Badass

Participants: Odalric Maillard, Emilie Kaufmann.

- *Title*:
- *Type*: National Research Agency
- *Coordinator*: Inria Lille (O. Maillard)
- *Duration*: 2016-2020
- *Abstract*: Motivated by the fact that a number of modern applications of sequential decision making require developing strategies that are especially robust to change in the stationarity of the signal, and in order to anticipate and impact the next generation of applications of the field, the BADASS project intends to push theory and application of MAB to the next level by incorporating non-stationary observations while retaining near optimality against the best not necessarily constant decision strategy. Since a non-stationary process typically decomposes into chunks associated with some possibly hidden variables (states), each corresponding to a stationary process, handling non-stationarity crucially requires exploiting the (possibly hidden) structure of the decision problem. For the same reason, a MAB for which arms can be arbitrary non-stationary processes is powerful enough to capture MDPs and even partially observable MDPs as special cases, and it is thus important to jointly address the issue of non-stationarity together with that of structure. In order to advance these two nested challenges from a solid theoretical standpoint, we intend to focus on the following objectives: *(i)* To broaden the range of optimal strategies for stationary MABs: current strategies are only known to be provably optimal in a limited range of scenarios for which the class of distribution (structure) is perfectly known; also, recent heuristics possibly adaptive to the class need to be further analyzed. *(ii)* To strengthen the literature on pure sequential prediction (focusing on a single arm) for non-stationary signals via the construction of adaptive confidence sets and a novel measure of complexity: traditional approaches consider a worst-case scenario and are thus overly conservative and non-adaptive to simpler signals. *(iii)* To embed the low-rank matrix completion and spectral methods in the context of reinforcement learning, and further study models of structured environments: promising heuristics in the context of e.g. contextual MABs or Predictive State Representations require stronger theoretical guarantees.

  This project will result in the development of a novel generation of strategies to handle non-stationarity and structure that will be evaluated in a number of test beds and validated by a rigorous theoretical analysis. Beyond the significant advancement of the state of the art in MAB and RL theory and the mathematical value of the program, this JCJC BADASS is expected to strategically impact societal and industrial applications, ranging from personalized health-care and e-learning to computational sustainability or rain-adaptive river-bank management to cite a few.

9.1.3. ANR ExTra-Learn

Participants: Alessandro Lazaric, Jérémie Mary, Rémi Munos, Michal Valko.

- *Title*: Extraction and Transfer of Knowledge in Reinforcement Learning
- *Type*: National Research Agency (ANR-9011)
- *Coordinator*: Inria Lille (A. Lazaric)
- *Duration*: 2014-2018
- *Abstract*: ExTra-Learn is directly motivated by the evidence that one of the key features that allows humans to accomplish complicated tasks is their ability of building knowledge from past experience and transfer it while learning new tasks. We believe that integrating transfer of learning in machine learning algorithms will dramatically improve their learning performance and enable them to solve complex tasks. We identify in the reinforcement learning (RL) framework the most suitable candidate for this integration. RL formalizes the problem of learning an optimal control policy from the experience directly collected from an unknown environment. Nonetheless, practical limitations of current algorithms encouraged research to focus on how to integrate prior knowledge

into the learning process. Although this improves the performance of RL algorithms, it dramatically reduces their autonomy. In this project we pursue a paradigm shift from designing RL algorithms incorporating prior knowledge, to methods able to incrementally discover, construct, and transfer "prior" knowledge in a fully automatic way. More in detail, three main elements of RL algorithms would significantly benefit from transfer of knowledge. *(i)* For every new task, RL algorithms need exploring the environment for a long time, and this corresponds to slow learning processes for large environments. Transfer learning would enable RL algorithms to dramatically reduce the exploration of each new task by exploiting its resemblance with tasks solved in the past. *(ii)* RL algorithms evaluate the quality of a policy by computing its state-value function. Whenever the number of states is too large, approximation is needed. Since approximation may cause instability, designing suitable approximation schemes is particularly critical. While this is currently done by a domain expert, we propose to perform this step automatically by constructing features that incrementally adapt to the tasks encountered over time. This would significantly reduce human supervision and increase the accuracy and stability of RL algorithms across different tasks. *(iii)* In order to deal with complex environments, hierarchical RL solutions have been proposed, where state representations and policies are organized over a hierarchy of subtasks. This requires a careful definition of the hierarchy, which, if not properly constructed, may lead to very poor learning performance. The ambitious goal of transfer learning is to automatically construct a hierarchy of skills, which can be effectively reused over a wide range of similar tasks.

- *Activity Report*: Research in ExTra-Learn continued in investigating how knowledge can be transferred into reinforcement learning algorithms to improve their performance. Pierre-Victor Chaumier did a 4 months internship in SequeL studying how to perform transfer neural networks across different games in the Atari platform. Unfortunately, the preliminary results we obtained were not very positive. We investigated different transfer models, from basic transfer of a fully trained network, to co-train over multiple games and retrain with initialization from a previous network. In most of the cases, the improvement from transfer was rather limited and in some cases even negative transfer effects appeared. This seems to be intrinsic in the neural network architecture which tends to overfit on one single task and it poorly generlizes over alternative tasks. Another activity was related to the study of macro-actions in RL. We proved for the first time under which conditions macro-actions can actually improve the learning speed of an RL exploration-exploitation algorithm. This is the first step towards the automatic identification and construction of useful macro-actions across multiple tasks.

9.1.4. ANR KEHATH

Participant: Olivier Pietquin.

- *Acronym*: KEHATH
- *Title*: Advanced Quality Methods for Post-Edition of Machine Translation
- *Type*: ANR
- *Coordinator*: Lingua & Machina
- *Duration*: 2014-2017
- *Other partners*: Univ. Lille 1, Laboratoire d'Informatique de Grenoble (LIG)
- *Abstract*: The translation community has seen a major change over the last five years. Thanks to progress in the training of statistical machine translation engines on corpora of existing translations, machine translation has become good enough so that it has become advantageous for translators to post-edit machine outputs rather than translate from scratch. However, current enhancement of machine translation (MT) systems from human post-edition (PE) are rather basic: the post-edited output is added to the training corpus and the translation model and language model are re-trained, with no clear view of how much has been improved and how much is left to be improved. Moreover, the final PE result is the only feedback used: available technologies do not take advantages of logged sequences of post-edition actions, which inform on the cognitive processes of the post-editor. The

KEHATH project intends to address these issues in two ways. Firstly, we will optimise advanced machine learning techniques in the MT+PE loop. Our goal is to boost the impact of PE, that is, reach the same performance with less PE or better performance with the same amount of PE. In other words, we want to improve machine translation learning curves. For this purpose, active learning and reinforcement learning techniques will be proposed and evaluated. Along with this, we will have to face challenges such as MT systems heterogeneity (statistical and/or rule-based), and ML scalability so as to improve domain-specific MT. Secondly, since quality prediction (QP) on MT outputs is crucial for translation project managers, we will implement and evaluate in real-world conditions several confidence estimation and error detection techniques previously developed at a laboratory scale. A shared concern will be to work on continuous domain-specific data flows to improve both MT and the performance of indicators for quality prediction. The overall goal of the KEHATH project is straightforward: gain additional machine translation performance as fast as possible in each and every new industrial translation project, so that post-edition time and cost is drastically reduced. Basic research is the best way to reach this goal, for an industrial impact that is powerful and immediate.

9.1.5. ANR MaRDi

Participants: Olivier Pietquin, Bilal Piot.

- *Acronym*: MaRDi
- *Title*: Man-Robot Dialogue
- *Type*: ANR
- *Coordinator*: Univ. Lille 1 (Olivier Pietquin)
- *Duration*: 2012-2016
- *Other partners*: Laboratoire d'Informatique d'Avignon (LIA), CNRS - LAAS (Toulouse), Acapela group (Toulouse)
- *Abstract*: In the MaRDi project, we study the interaction between humans and machines as a situated problem in which human users and machines share the same environment. Especially, we investigate how the physical environment of robots interacting with humans can be used to improve the performance of spoken interaction which is known to be imperfect and sensible to noise. To achieve this objectif, we study three main problems. First, how to interactively build a multimodal representation of the current dialogue context from perception and proprioception signals. Second, how to automatically learn a strategy of interaction using methods such as reinforcement learning. Third, how to provide expressive feedbacks to users about how the machine is confident about its behaviour and to reflect its current state (also the physical state).

9.1.6. National Partners

- CentraleSupélec
  - J.Perolat, B.Piot and O.Pietquin worked with M.Geist on Stochastic Games. it led to a conference publication in ICML 2016.
- Inria Nancy - Grand Est
  - J.Perolat, B.Piot and O.Pietquin worked with Bruno Scherrer on Stochastic Games. It led to a conference publication in AISTATS 2016 [47] and ICML 2016.
- Institut de Mathématiques de Toulouse
  - É. Kaufmann had publications at COLT, ALT and NIPS with Aurélie Garivier.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

Program: H2020

Project acronym: BabyRobot

Project title: Child-Robot Communication and Collaboration

Duration: 01/2016 - 12/2018

Coordinator: Alexandros Potamianos (Athena Research and Innovation Center in Information Communication and Knowledge Technologies, Greece)

Other partners: Institute of Communication and Computer Systems (Greece), The University of Hertfordshire Higher Education Corporation (UK), Universitaet Bielefeld (Germany), Kunlgliga Tekniska Hoegskolan (Sweden), Blue Ocean Robotics ApS (Denmark), Univ. Lille (France), Furhat Robotics AB (Sweden)

Abstract: The crowning achievement of human communication is our unique ability to share intentionality, create and execute on joint plans. Using this paradigm we model human-robot communication as a three step process: sharing attention, establishing common ground and forming shared goals. Prerequisites for successful communication are being able to decode the cognitive state of people around us (mindreading) and building trust. Our main goal is to create robots that analyze and track human behavior over time in the context of their surroundings (situational) using audio-visual monitoring in order to establish common ground and mind-reading capabilities. On BabyRobot we focus on the typically developing and autistic spectrum children user population. Children have unique communication skills, are quick and adaptive learners, eager to embrace new robotic technologies. This is especially relevant for special eduation where the development of social skills is delayed or never fully develops without intervention or therapy. Thus our second goal is to define, implement and evaluate child-robot interaction application scenarios for developing specific socio-affective, communication and collaboration skills in typically developing and autistic spectrum children. We will support not supplant the therapist or educator, working hand-inhand to create a low risk environment for learning and cognitive development. Breakthroughs in core robotic technologies are needed to support this research mainly in the areas of motion planning and control in constrained spaces, gestural kinematics, sensorimotor learning and adaptation. Our third goal is to push beyond the state-of-the-art in core robotic technologies to support natural human-robot interaction and collaboration for edutainment and healthcare applications. Creating robots that can establish communication protocols and form collaboration plans on the fly will have impact beyond the application scenarios investigated here.

9.2.2. Collaborations in European Programs, Except FP7 & H2020

Program: CHIST-ERA

Project acronym: IGLU

Project title: Interactively Grounded Language Understanding

Duration: 11/2015 - 10/2018

Coordinator: Jean Rouat (Université de Sherbrooke, Canada)

Other partners: UMONS (Belgique), Inria (France), Univ-Lille (France), KTH (sweden), Universidad de Zaragoza (Spain)

Abstract: Language is an ability that develops in young children through joint interaction with their caretakers and their physical environment. At this level, human language understanding could be referred as interpreting and expressing semantic concepts (e.g. objects, actions and relations) through what can be perceived (or inferred) from current context in the environment. Previous work in the field of artificial intelligence has failed to address the acquisition of such perceptually-grounded knowledge in virtual agents (avatars), mainly because of the lack of physical embodiment (ability to interact physically) and dialogue, communication skills (ability to interact verbally). We believe that robotic agents are more appropriate for this task, and that interaction is a so important aspect of human language learning and understanding that pragmatic knowledge (identifying or conveying intention) must be present to complement semantic knowledge. Through a developmental approach where knowledge grows in complexity while driven by multimodal experience and language interaction with a human, we propose an agent that will incorporate models of dialogues, human emotions and intentions as part of its decision-making process. This will lead anticipation and reaction not only based on its internal state (own goal and intention, perception of the environment), but also on the perceived state and intention of the human interactant. This will be possible through the development of advanced machine learning methods (combining developmental, deep and reinforcement learning) to handle large-scale multimodal inputs, besides leveraging state-of-the-art technological components involved in a language-based dialog system available within the consortium. Evaluations of learned skills and knowledge will be performed using an integrated architecture in a culinary use-case, and novel databases enabling research in grounded human language understanding will be released.

9.3. International Initiatives

9.3.1. Inria Associate Teams Not Involved in an Inria International Labs

9.3.1.1. EduBand

Title: Educational Bandits

International Partner (Institution - Laboratory - Researcher):

Carnegie Mellon University (United States) - Department of Computer Science, Theory of computation lab - Emma Brunskill

Start year: 2015

See also: https://project.inria.fr/eduband/

Education can transform an individual's capacity and the opportunities available to him. The proposed collaboration will build on and develop novel machine learning approaches towards enhancing (human) learning. Massive open online classes (MOOCs) are enabling many more people to access education, but mostly operate using status quo teaching methods. Even more important than access is the opportunity for online software to radically improve the efficiency, engagement and effectiveness of education. Existing intelligent tutoring systems (ITSs) have had some promising successes, but mostly rely on learning sciences research to construct hand-built strategies for automated teaching. Online systems make it possible to actively collect substantial amount of data about how people learn, and offer a huge opportunity to substantially accelerate progress in improving education. An essential aspect of teaching is providing the right learning experience for the student, but it is often unknown a priori exactly how this should be achieved. This challenge can often be cast as an instance of decision-making under uncertainty. In particular, prior work by Brunskill and colleagues demonstrated that reinforcement learning (RL) and multi-arm bandit (MAB) can be very effective approaches to solve the problem of automated teaching. The proposed collaboration is thus intended to explore the potential interactions of the fields of online education and RL and MAB. On the one hand, we will define novel RL and MAB settings and problems in online education. On the other hand, we will investigate how solutions developed in RL and MAB could be integrated in ITS and MOOCs and improve their effectiveness.

9.3.2. Inria International Partners

9.3.2.1. With CWI

Title: Learning theory

"North-European Associate Team"

Centrum Wiskunde & Informatica (CWI), Amsterdam (NL) - Peter Grünwald

Duration: 2016 - 2018

Start year: 2016

ABSTRACT: The aim is to develop the theory of learning for sequential decision making under uncertainty problems.

In 2016, this collaboration involved D. Ryabko, É. Kaufmann, J. Ridgway, M. Valko, A. Lazaric, O. Maillard. A post-doc funded by Inria has been recruited in Fall 2016.

This collaboration aims at developing through the Inria International Laboratory with CWI.

9.3.2.2. With University of Leoben

Title: The multi-armed bandit problem

International Partner (Institution - Laboratory - Researcher):

University of Leoben (Austria) - Peter Auer

Duration: 2016 - 2016

Start year: 2016

ABSTRACT: Study of the multi-armed bandit problem.

9.3.2.3. Informal International Partners

- University of California Irvine (USA)

  Anima Anandkumar *Collaborator*

  A. Lazaric collaborates with A. Anandkumar on the use of spectral methods for reinforcement learning.

- University of Lancaster (UK)

  Borja Balle *Collaborator*

  O-A. Maillard collaborates with B. Balle on concentration inequalities for Hankel matrices.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

9.4.1.1. Internships

- Cricia Zilda Felicio Paixao, University Uberlandia, Brasil, Sep. 2015-Jul. 2016, working on recommendation systems in collaboration with Philippe Preux
- Maryam Aziz, Northeastern University, May-Aug. 2016, working on multi-armed bandits for clinical trials in collaboration with Emilie Kaufmann
- Kamyar Azizzadenesheli, University of California at Irvine, Aug-Oct. 2016, working on latent variable models for reinforcement learning in collaboration with Alessandro Lazaric
- Pierre-Victor Chaumier, Ecole Polytechnique, Jan-Jun. 2016, working on transfer learning in collaboration with Alessandro Lazaric
- Firas Jarboui, ENSTA ParisTech, France, May-July.@ 2016, working on Human-AI co-operation, in collaboration with Christos Dimitrakakis.

9.4.2. Visits to International Teams

9.4.2.1. Research Stays Abroad

- Christos Dimitrakakis visited SEAS, Harvard University, USA in the context of a Swedish/EU project “Market Mechanisms for Multiple Minds”, and the future of life institute project “Mechanism Design for Multiple AIs”, May-June, September-December 2016.
- Christos Dimitrakakis visited ETHZ, Switzerland, in the context of the Swiss SNSF project “Differential Privacy and Approximate Decision Making”, July-September 2016.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- C. Dimitrakakis, ICML Workshop on the theory and practice of differential privacy.
- Ph. Preux, "Big Data : Modelisation, Estimation and Selection", June 2016, Villeneuve d'Ascq.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Conference on Learning Theory (COLT)
- International Joint Conference on Artificial Intelligence (IJCAI)
- European Conference on Machine Learning (ECML)
- ICPRAM
- French conferences:
  - Extraction et Gestion de Conaissances (EGC),
  - Journées Francophones de Planification, Décision, Apprentissage (JFPDA),
  - Apprentissage Automatique et Fouille de Données & Société Française de Classification

10.1.2.2. Reviewer

- Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16)
- Conference on Learning Theory (COLT 2016)
- European Workshop on Reinforcement Learning (EWRL 2016)
- European Conference on Machine Learning (ECML 2016)
- International Conference on Machine Learning (ICML 2016)
- Neural Information Processing Systems (NIPS 2016)
- International Joint Conference on Artificial Intelligence (IJCAI 2016)
- Conference on Autonomous Agents and Multia-Agent Systems (AAMAS 2016)
- International Conference on Artificial Intelligence and Statistics (AISTATS 2016)
- French conferences:
  - Extraction et Gestion de Conaissances (EGC),
  - Journées Francophones de Planification, Décision, Apprentissage (JFPDA),
  - conférence francophone sur l'Apprentissage Automatique (CAp),
  - Apprentissage Automatique et Fouille de Données & Société Française de Classification
  - Conférence Nationale d'Intelligence Artificielle (CNIA)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Journal of Games.
- Neurocomputing.
- Revue d'Intelligence Artificielle

10.1.3.2. Reviewer - Reviewing Activities

- Automatica

- Artificial Intelligence Journal
- Machine Learning Journal
- Journal of Artificial Intelligence Research
- Journal of Machine Learning Research
- AMS Mathematical Review
- IEEE Transaction on Signal Processing
- IEEE Tansaction on Cybernetics

10.1.4. Invited Talks

- R. Gaudel, *From Bandits to Recommender Systems*, Presented on September 29th, 2016 at ENSAI, Rennes, France
- R. Gaudel, *Recommendation as a sequential process*, Presented on December 12th, 2016 at CMLA Mathématiques Appliquées, Cachan, France
- E. Kaufmann, *The information complexity of best arm identification*, Multi-armed Bandit Workshop 2016 at STOR-i, Lancaster University, UK, January 2016.
- E. Kaufmann, *The information complexity of sequential resource allocation*, seminar of the Collegio Carlo Alberto, Turin, March 2016.
- E. Kaufmann, *Optimal Best Arm Identification with Fixed Confidence*, Workshop on Computational and Statistical Trade-offs in Learning , Institut des Hautes Etudes Scientifiques (Orsay), March 2016.
- E. Kaufmann, *The information complexity of sequential resource allocation*, Stalab seminar, University of Cambridge, UK, April 2016.
- E. Kaufmann, *Stratégies bayésiennes et fréquentistes dans un modèle de bandit*, 1er congrès de la Société Mathématique de France, Tours, June 2016.
- E. Kaufmann, *Stratégies bayésiennes et fréquentistes dans un modèle de bandit*, Journées MAS, Grenoble, August 2016.
- E. Kaufmann, *Revisiting the Exploration-Exploitation Tradeoff in Bandit Models*, Workshop on Optimization and Decision-Making in Uncertainty, Simons Institute, Berkeley, September 2016.
- A. Lazaric, *Spectral Methods for Learning in POMDPs*, University of Liège, Belgium, February 2016.
- A. Lazaric, *Spectral Methods for Learning in POMDPs*, CMLA Mathématiques Appliquées, Cachan, France, February 2016.
- A. Lazaric, *Incremental Kernel Regression with Ridge Leverage Score Sampling*, “Data Learning and Inference” (DALI), Sestri Levante, Italy, April 2016.
- A. Lazaric, *Optimism and Randomness in Linear Multi-armed Bandit*, “International Conference on Monte-Carlo Techniques”, July 2016.
- J. Mary, *Structured Bandits*, “University of Strasbourg”, May. 2016.
- J. Mary, *Tutorial on Deep Neural Networks*, “Journées Big Data”, by the Laboratoire Painlevé. Jun. 2016.
- J. Mary, *Machine Learning and AI*, “EDF Seminar”, Dec. 2016.
- O. Pietquin, *Closing the Interaction Loop with (Inverse) Reinforcement Learning*, Presented on November 15, 2016 at AWRL, Hamilton, New-Zealand
- O. Pietquin, *Challenges of End-to-End Spoken Dialogue Systems*, Presented on December 10, 2016 at FILM@NIPS Workshop, Barcelona, Spain
- O. Pietquin, *Keeping the Human in the Loop: Challenges for Machine Learning*, Presented on March 10, 2016 at Xerox Research Center in Europe, Grenoble, France

- M. Valko, *Spectral Methods for Learning in POMDPs*, University of Liège, Belgium, February 2016.
- M. Valko, *Where is Justin Bieber?*, Presented on September 22nd, 2016 at Comenius University in Bratislava, Slovakia (*FMFI 2016*)
- M. Valko, *Bandit learning*, Presented on September 15–19th, 2016 at Information technologies - Applications and Theory, at Tatranské Matliare, High Tatras, Slovakia (*ITAT 2016*)
- M. Valko, *Decision-making on graphs without graphs*, Presented on June 16-17th, 2016 at Graph-based Learning and Graph Mining workshop, at Inria Lille, France (*GBLGM 2016*)
- M. Valko, *Sequential learning on graphs with limited feedback*, Presented on May 11–13th, 2016 at Data Driven Approach to Networks and Language, at ENS Lyon, France (*NETSpringLyon 2016*)
- M. Valko, *Benefits of Graphs in Bandit Settings*, Presented on January 11–12th, 2016 at Multi-armed Bandit Workshop 2016 at STOR-i, Lancaster University, UK (*STOR-i 2016*)

10.1.5. Scientific Expertise

- Agence Nationale pour la Recherche (ANR)
- ANRT
- D2RT Ile de France
- Institut National de Recherche en Agronomie (INRA)
- Fonds National pour la Recherche Scientifique (FNRS), Belgium
- H2020 program
- *A. Lazaric* was a member of the hiring committee for junior researchers at Inria Lille (2016).
- *M. Valko* is an elected member of the evaluation committee and participates in the hiring, promotion, and evaluation juries of Inria, notably
  - Hiring committee for junior researchers at Inria Sophia Antipolis (2016)
  - Selection committee for Inria award for scientific excellence, junior and senior (2016)
  - Selection committee for CR promotions (2016)
- Ph. Preux has chaired the hiring committee for an associate professor position at Université de Lille 3
- J. Mary was webpage chair for ICML'2016 in NYC

10.1.6. Research Administration

- Philippe Preux is:
  - Délégué Scientifique Adjoint (DSA) at Inria Lille
  - member of the Evaluation Committee (CE) at Inria
  - member of the Project Committee Board (BCP, Bureau du Comité des Projets) at Inria Lille
  - head of the “Data Intelligence” (DatInG) thematic group at CRIStAL.
  - member of the Scientific Committee of CRIStAL.
- R. Gaudel is member of the board of CRIStAL.
- R. Gaudel is manager of proml mailing list. This mailing list gathers French-speaking researchers from Machine Learning community.
- J. Mary is member of the "Commission Développement Technologique" at Inria Lille.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence: C. Dimitrakakis, C2i, 25h eqTD, L1-2, Université de Lille 3, France.

Licence: C. Dimitrakakis, Traitement de données, Université de Lille 3, France.

Licence: C. Dimitrakakis, Modélisation de bases de données, Université de Lille 3, France.

Licence: C. Dimitrakakis, Fonctionnement des réseaux, Université de Lille 3, France.

Master: A. Lazaric, Reinforcement Learning, 25h eqTD, M2, ENS Cachan, France

Master: A. Lazaric, Reinforcement Learning, 25h eqTD, M2, Ecole Centrale Lille, France

Master: Ph. Preux, Advanced data mining, 30h eqTD, M2, Université de Lille 3, France

Master: Ph. Preux, Fundamental algorithms for data mining, 30h eqTD, M1, Université de Lille 3, France

Licence: Ph. Preux, Neural Networks, 28h eqTD, L3, Université de Lille 3, France

Licence: Ph. Preux, Graph Theory, 28h eqTD, L3, Université de Lille 3, France

Licence: Ph. Preux, C2i, 25h eqTD, L1-2, Université de Lille 3, France.

Master: M. Valko, 2016/2017 Fall: Graphs in Machine Learning, 27h eqTD, M2, ENS Cachan

Licence: R. Gaudel, 2016/2017 Spring: programmation R pour statistiques et sociologie quantitative, 44h eqTD, L1, université Lille 3, France

Licence: R. Gaudel, 2016/2017 Fall: préparation au C2i niveau 1, 30h eqTD, L1-3, université Lille 3, France

Licence: R. Gaudel, 2016/2017 Spring: préparation au C2i niveau 1, 25h eqTD, L1-3, université Lille 3, France

Licence: R. Gaudel, 2016/2017 Fall: travail collaboratif et à distance dans un monde numérique, 13h eqTD, L1-3 (enseignement à distance), université Lille 3, France

Licence: R. Gaudel, 2016/2017 Fall: algorithmes fondamentaux de la fouille de données, 30h eqTD, M1, université Lille 3, France

Licence: R. Gaudel, 2016/2017 Fall: Fouille de données avancée, 30h eqTD, M2, université Lille 3, France

Master: B. Piot, 2016/2017 Spring: Web Design, 60h eqTD, M2, Univ. Lille, France

Master: B. Piot, 2016/2017 Spring: Object Programming, 70h eqTD, M2, Univ. Lille, France

Master: B. Piot, 2016/2017 Fall: Web Design, 30h eqTD, M2, Univ. Lille, France

Master: B. Piot, 2016/2017 Fall: Object Programming, 22h eqTD, M2, Univ. Lille, France

Master: B. Piot, 2016/2017 Fall: Databases, 30h eqTD, M1, Univ. Lille, France

Master: J. Mary, 2016/2017 Fall: algorithmes fondamentaux de la fouille de données, 30h eqTD, M1, Univ. Lille, France

Master: J. Mary, 2016/2017 Fall: Programmation Web Avancée, 30h eqTD, M2, Univ. Lille, France

Master: J. Mary, 2016/2017 Fall: Reinforcement Learning, 16h eqTD, M2, Univ. Lille, France

10.2.2. Supervision

HdR: Michal Valko, Bandits on graphs and structures, ENS-Cachan, June 15th, 2016

PhD: Frédéric Guillou, On recommendation systems in sequential context, University of Lille, Dec. 2nd, 2016, advisors: Philippe Preux, Romaric Gaudel, Jérémie Mary

PhD: Tomas Kocak, Apprentissage séquentiel avec similitudes, University of Lille, Nov. 28th, 2016, advisor: Michal Valko, Rémi Munos

PhD: Hadrien Glaude, Méthodes des moments pour l'inférence de systèmes séquentiels linéaires rationnels, University of Lille, July. 8th, 2016, advisor: Olivier Pietquin

PhD in progress: Pratik Gajane, Sequential Learning and Decision Making under Partial Monitoring, University of Lille, started Oct. 2014, advisor: Philippe Preux

PhD in progress: Marc Abeille, Randomized Exploration-exploration Stretegies, University of Lille, started Oct. 2014, advisor: Alessandro Lazaric

PhD in progress: Merwan Barlier, Dialogues intelligents basés sur l'écoute de conversations homme/homme, University of Lille, started Oct. 2014, advisor: Olivier Pietquin

PhD in progress: Alexandre Berard, Learning from post-editing for machine translation, University of Lille, started Oct. 2014, advisor: Olivier Pietquin

PhD in progress: Lilian Besson, Apprentissage séquentiel multi-joueurs pour la radio intelligente, CentraleSupélec Rennes, started Oct. 2016, advisor: Emilie Kaufmann

PhD in progress: Reda Alami, Bandit à Mémoire pour la prise de décision en environnement dynamique, Orange LABS, University of Paris-Saclay, started Oct. 2016, advisor: Odalric-Ambrym Maillard, Raphaël Feraud

PhD in progress: Daniele Calandriello, Efficient Sequential Learning in Structured and Constrained Environment, Inria, started Oct. 2014, advisor: Michal Valko, Alessandro Lazaric

PhD in progress: Ronan Fruit, Transfer in Hierarchical Reinforcement Learning, University of Lille, started Dec. 2015, advisor: Alessandro Lazaric

PhD in progress: Guillaume Gautier, DPPs in ML, started Oct. 2016, advisor: Michal Valko; Rémi Bardenet

PhD in progress: Jean-Bastien Grill, Création et analyse d'algorithmes efficaces pour la prise de décision dans un environnement inconnu et incertain, Inria/ENS Paris/Lille 1, started Oct. 2014, advisor: Rémi Munos, Michal Valko

PhD in progress: Julien Perolat, Reinforcement learning: the 2-player case, University of Lille, started Oct. 2014, advisor: Olivier Pietquin, Bilal Piot

PhD in progress: Florian Strub, Deep sequential learning and its application to human-robot interaction, University of Lille, started Jan. 2016, advisor: Olivier Pietquin, Jérémie Mary

PhD in progress: Romain Warlop, Novel Learning and Exploration-Exploitation Methods for Effective Recommender Systems, University of Lille, started Sep. 2015, advisor: Jérémie Mary

PhD in progress: Aristide Tossou, Privacy in Sequential Decision Making (provional), Chalmers, started Feb. 2015, advisor: Christos Dimitrakakis

10.2.3. Juries

PhD and hdr juries:

- E. Kaufmann: Marie-Liesse Cauwet, LRI, Orsay.
- A. Lazaric: Matteo Pirotta, Politecnico di Milano, Italy.
- J. Mary: examinator for Raphël Puget, université Paris 6.
- J. Mary: reviewer for Robin Allesiardo, université Paris Saclay
- Ph. Preux: reviewer for Hongliang Zhong, Laboratoire d'Informatique Fondamentale, Marseille
- Ph. Preux: president of the defense jury of the HDR of Matthieu Geist, Université de Lille
- O. Pietquin: advisor of the HDR of Matthieu Gesit, Université de Lille
- O. Pietquin: Hatim Khouzami, University of Avignon

PhD mid-term evaluation:

- A. Lazaric: Claire Vernade (mid-term evaluation), Telecom ParisTech, France.
- E. Kaufmann: opponent for the licenciate thesis of Stefan Magureanu, KTH, Stockholm, Sweden.

10.3. Popularization

- A. Lazaric was interviewed for Inria-Lille magazine (December issue).
- J. Mary gave a 55 min talk in front of students of Lycée Sainte-Famille at Amiens.
- Ph. Preux gave 2 talks on “Artificial Intelligence” in a high-school in Villeneuve d’Ascq within the “Fête de la science”.
- O. Pietquin was interviewed by France Culture (Supersonic) on March 23rd, 2016

11. Bibliography

Major publications by the team in recent years

[1] O. Cappé, A. Garivier, O.-A. Maillard, R. Munos, G. Stoltz.*Kullback-Leibler Upper Confidence Bounds for Optimal Sequential Allocation*, in "Annals of Statistics", 2013, vol. 41, n^{o} 3, p. 1516-1541, Accepted, to appear in Annals of Statistics, https://hal.archives-ouvertes.fr/hal-00738209.

[2] A. Carpentier, M. Valko.*Revealing graph bandits for maximizing local influence*, in "International Conference on Artificial Intelligence and Statistics", Seville, Spain, May 2016, https://hal.inria.fr/hal-01304020.

[3] N. Gatti, A. Lazaric, M. Rocco, F. Trovò.*Truthful Learning Mechanisms for Multi–Slot Sponsored Search Auctions with Externalities*, in "Artificial Intelligence", October 2015, vol. 227, p. 93-139, https://hal.inria.fr/hal-01237670.

[4] M. Ghavamzadeh, Y. Engel, M. Valko.*Bayesian Policy Gradient and Actor-Critic Algorithms*, in "Journal of Machine Learning Research", January 2016, vol. 17, n^{o} 66, p. 1-53, https://hal.inria.fr/hal-00776608.

[5] H. Kadri, E. Duflos, P. Preux, S. Canu, A. Rakotomamonjy, J. Audiffren.*Operator-valued Kernels for Learning from Functional Response Data*, in "Journal of Machine Learning Research (JMLR)", 2016, https://hal.archives-ouvertes.fr/hal-01221329.

[6] E. Kaufmann, O. Cappé, A. Garivier.*On the Complexity of Best Arm Identification in Multi-Armed Bandit Models*, in "Journal of Machine Learning Research", January 2016, vol. 17, p. 1-42, https://hal.archives-ouvertes.fr/hal-01024894.

[7] A. Lazaric, M. Ghavamzadeh, R. Munos.*Analysis of Classification-based Policy Iteration Algorithms*, in "Journal of Machine Learning Research", 2016, vol. 17, p. 1 - 30, https://hal.inria.fr/hal-01401513.

[8] R. Munos.*From Bandits to Monte-Carlo Tree Search: The Optimistic Principle Applied to Optimization and Planning*, 2014, 130 pages, https://hal.archives-ouvertes.fr/hal-00747575.

[9] R. Ortner, D. Ryabko, P. Auer, R. Munos.*Regret bounds for restless Markov bandits*, in "Journal of Theoretical Computer Science (TCS)", 2014, vol. 558, p. 62-76 [*DOI :* 10.1016/j.tcs.2014.09.026], https://hal.inria.fr/hal-01074077.

[10] D. Ryabko, J. Mary.*A Binary-Classification-Based Metric between Time-Series Distributions and Its Use in Statistical and Learning Problems*, in "Journal of Machine Learning Research", 2013, vol. 14, p. 2837-2856, https://hal.inria.fr/hal-00913240.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] H. GLAUDE.*Learning rational linear sequential systems using the method of moments*, Université de Lille 1 - Sciences et Technologies, July 2016, https://tel.archives-ouvertes.fr/tel-01374080.

[12] F. GUILLOU.*On Recommendation Systems in a Sequential Context*, Université Lille 3, December 2016, https://tel.archives-ouvertes.fr/tel-01407336.

[13] V. MUSCO.*Propagation Analysis based on Software Graphs and Synthetic Data*, Université Lille 3, November 2016, https://tel.archives-ouvertes.fr/tel-01398903.

[14] M. VALKO.*Bandits on graphs and structures*, École normale supérieure de Cachan - ENS Cachan, June 2016, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01359757.

Articles in International Peer-Reviewed Journal

[15] M. GHAVAMZADEH, Y. ENGEL, M. VALKO.*Bayesian Policy Gradient and Actor-Critic Algorithms*, in "Journal of Machine Learning Research", January 2016, vol. 17, n^{o} 66, p. 1-53, https://hal.inria.fr/hal-00776608.

[16] H. KADRI, E. DUFLOS, P. PREUX, S. CANU, A. RAKOTOMAMONJY, J. AUDIFFREN.*Operator-valued Kernels for Learning from Functional Response Data*, in "Journal of Machine Learning Research (JMLR)", 2016, https://hal.archives-ouvertes.fr/hal-01221329.

[17] E. KAUFMANN, O. CAPPÉ, A. GARIVIER.*On the Complexity of Best Arm Identification in Multi-Armed Bandit Models*, in "Journal of Machine Learning Research", January 2016, vol. 17, p. 1-42, https://hal.archives-ouvertes.fr/hal-01024894.

[18] A. KHALEGHI, D. RYABKO.*Nonparametric multiple change point estimation in highly dependent time series*, in "Theoretical Computer Science", 2016, vol. 620, p. 119-133 [*DOI :* 10.1016/J.TCS.2015.10.041], https://hal.inria.fr/hal-01235330.

[19] A. KHALEGHI, D. RYABKO, J. MARY, P. PREUX.*Consistent Algorithms for Clustering Time Series*, in "Journal of Machine Learning Research", 2016, vol. 17, n^{o} 3, p. 1 - 32, https://hal.inria.fr/hal-01399613.

[20] A. LAZARIC, M. GHAVAMZADEH, R. MUNOS.*Analysis of Classification-based Policy Iteration Algorithms*, in "Journal of Machine Learning Research", 2016, vol. 17, p. 1 - 30, https://hal.inria.fr/hal-01401513.

[21] V. MUSCO, M. MONPERRUS, P. PREUX.*A Large-scale Study of Call Graph-based Impact Prediction using Mutation Testing*, in "Software Quality Journal", 2016 [*DOI :* 10.1007/S11219-016-9332-8], https://hal.inria.fr/hal-01346046.

[22] G. NEU, B. GÁBOR.*Importance Weighting Without Importance Weights: An Efficient Algorithm for Combinatorial Semi-Bandits*, in "Journal of Machine Learning Research", August 2016, vol. 17, n^{o} 154, p. 1 - 21, https://hal.archives-ouvertes.fr/hal-01380278.

International Conferences with Proceedings

[23] K. Azizzadenesheli, A. Lazaric, A. Anandkumar.*Reinforcement Learning of POMDPs using Spectral Methods*, in "Proceedings of the 29th Annual Conference on Learning Theory (COLT2016)", New York City, United States, June 2016, https://hal.inria.fr/hal-01322207.

[24] M. Barlier, R. Laroche, O. Pietquin.*A Stochastic Model for Computer-Aided Human-Human Dialogue*, in "Interspeech 2016", San Francisco, United States, September 2016, vol. 2016, p. 2051 - 2055, https://hal.inria.fr/hal-01406894.

[25] M. Barlier, R. Laroche, O. Pietquin.*Learning Dialogue Dynamics with the Method of Moments*, in "Workshop on Spoken Language Technologie (SLT 2016)", San Diego, United States, December 2016, https://hal.inria.fr/hal-01406904.

[26] D. Calandriello, A. Lazaric, M. Valko.*Analysis of Nyström method with sequential ridge leverage score sampling*, in "Uncertainty in Artificial Intelligence", New York City, United States, June 2016, https://hal.inria.fr/hal-01343674.

[27] A. Carpentier, M. Valko.*Revealing graph bandits for maximizing local influence*, in "International Conference on Artificial Intelligence and Statistics", Seville, Spain, May 2016, https://hal.inria.fr/hal-01304020.

[28] L. El Asri, R. Laroche, O. Pietquin.*Compact and Interpretable Dialogue State Representation with Genetic Sparse Distributed Memory*, in "7th International Workshop on Spoken Dialogue Systems (IWSDS 2016)", Saariselka, Finland, January 2016, https://hal.inria.fr/hal-01406873.

[29] L. El Asri, B. Piot, M. Geist, R. Laroche, O. Pietquin.*Score-based Inverse Reinforcement Learning*, in "International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2016)", Singapore, Singapore, May 2016, https://hal.inria.fr/hal-01406886.

[30] A. Erraqabi, M. Valko, A. Carpentier, O.-A. Maillard.*Pliable rejection sampling*, in "International Conference on Machine Learning", New York City, United States, June 2016, https://hal.inria.fr/hal-01322168.

[31] C. Z. Felício, K. V. R. Paixão, C. A. Z. Barcelos, P. Preux.*Preference-like Score to Cope with Cold-Start User in Recommender Systems*, in "28th International Conference on Tools with Artificial Intelligence (ICTAI)", San Jose, United States, Proceedings of the IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), November 2016, https://hal.inria.fr/hal-01390762.

[32] V. Gabillon, A. Lazaric, M. Ghavamzadeh, R. Ortner, P. Bartlett.*Improved Learning Complexity in Combinatorial Pure Exploration Bandits*, in "Proceedings of the 19th International Conference on Artificial Intelligence (AISTATS)", Cadiz, Spain, May 2016, https://hal.inria.fr/hal-01322198.

[33] A. Garivier, E. Kaufmann.*Optimal Best Arm Identification with Fixed Confidence*, in "29th Annual Conference on Learning Theory (COLT)", New York, United States, JMLR Workshop and Conference Proceedings, June 2016, vol. 49, https://hal.archives-ouvertes.fr/hal-01273838.

[34] A. Garivier, E. Kaufmann, W. M. Koolen.*Maximin Action Identification: A New Bandit Framework for Games*, in "29th Annual Conference on Learning Theory (COLT)", New-York, United States, JMLR Workshop and Conference Proceedings, June 2016, vol. 49, https://hal.archives-ouvertes.fr/hal-01273842.

[35] A. GARIVIER, E. KAUFMANN, T. LATTIMORE.*On Explore-Then-Commit Strategies*, in "NIPS", Barcelona, Spain, Advances in Neural Information Processing Systems (NIPS), December 2016, vol. 29, https://hal.archives-ouvertes.fr/hal-01322906.

[36] H. GLAUDE, O. PIETQUIN.*PAC learning of Probabilistic Automaton based on the Method of Moments*, in "International Conference on Machine Learning (ICML 2016)", New York, United States, June 2016, https://hal.inria.fr/hal-01406889.

[37] J.-B. GRILL, M. VALKO, R. MUNOS.*Blazing the trails before beating the path: Sample-efficient Monte-Carlo planning*, in "NIPS 2016 - Thirtieth Annual Conference on Neural Information Processing Systems", Barcelona, Spain, December 2016, https://hal.inria.fr/hal-01389107.

[38] F. GUILLOU, R. GAUDEL, P. PREUX.*Large-scale Bandit Recommender System*, in "the 2nd International Workshop on Machine Learning, Optimization and Big Data (MOD'16)", Volterra, Italy, August 2016, https://hal.inria.fr/hal-01406389.

[39] F. GUILLOU, R. GAUDEL, P. PREUX.*Scalable explore-exploit Collaborative Filtering*, in "Pacific Asia Conference on Information Systems (PACIS'16)", Chiayi, Taiwan, 2016, https://hal.inria.fr/hal-01406418.

[40] F. GUILLOU, R. GAUDEL, P. PREUX.*Sequential Collaborative Ranking Using (No-)Click Implicit Feedback*, in "The 23rd International Conference on Neural Information Processing (ICONIP'16)", Kyoto, Japan, Lecture Notes in Computer Science, October 2016, vol. 9948, p. 288 - 296 [*DOI :* 10.1007/978-3-319-46672-9_33], https://hal.inria.fr/hal-01406338.

[41] E. KAUFMANN, T. BONALD, M. LELARGE.*A Spectral Algorithm with Additive Clustering for the Recovery of Overlapping Communities in Networks*, in "ALT 2016 - Algorithmic Learning Theory", Bari, Italy, R. ORTNER, H. U. SIMON, S. ZILLES (editors), Lecture Notes in Computer Science, Springer, October 2016, vol. 9925, p. 355-370 [*DOI :* 10.1007/978-3-319-46379-7_24], https://hal.archives-ouvertes.fr/hal-01163147.

[42] T. KOCÁK, G. NEU, M. VALKO.*Online learning with Erdős-Rényi side-observation graphs*, in "Uncertainty in Artificial Intelligence", New York City, United States, June 2016, https://hal.inria.fr/hal-01320588.

[43] T. KOCÁK, G. NEU, M. VALKO.*Online learning with noisy side observations*, in "International Conference on Artificial Intelligence and Statistics", Seville, Spain, May 2016, https://hal.inria.fr/hal-01303377.

[44] V. MUSCO, A. CARETTE, M. MONPERRUS, P. PREUX.*A Learning Algorithm for Change Impact Prediction*, in "5th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering", Austin, United States, May 2016, https://hal.inria.fr/hal-01279620.

[45] V. MUSCO, M. MONPERRUS, P. PREUX.*Mutation-Based Graph Inference for Fault Localization*, in "International Working Conference on Source Code Analysis and Manipulation", Raleigh, United States, October 2016, https://hal.inria.fr/hal-01350515.

[46] J. PÉROLAT, B. PIOT, M. GEIST, B. SCHERRER, O. PIETQUIN.*Softened Approximate Policy Iteration for Markov Games*, in "ICML 2016 - 33rd International Conference on Machine Learning", New York City, United States, June 2016, https://hal.inria.fr/hal-01393328.

[47] J. PÉROLAT, B. PIOT, B. SCHERRER, O. PIETQUIN.*On the Use of Non-Stationary Strategies for Solving Two-Player Zero-Sum Markov Games*, in "19th International Conference on Artificial Intelligence and Statistics (AISTATS 2016)", Cadiz, Spain, Proceedings of the International Conference on Artificial Intelligences and Statistics, May 2016, https://hal.inria.fr/hal-01291495.

[48] D. RYABKO.*Things Bayes can't do*, in "Proceedings of the 27th International Conference on Algorithmic Learning Theory (ALT'16)", Bari, Italy, October 2016, vol. LNCS, n° 9925, p. 253-260 [*DOI :* 10.1007/978-3-319-46379-7_17], https://hal.inria.fr/hal-01380063.

[49] F. STRUB, R. GAUDEL, J. MARY.*Hybrid Recommender System based on Autoencoders*, in "the 1st Workshop on Deep Learning for Recommender Systems", Boston, United States, September 2016, p. 11 - 16 [*DOI :* 10.1145/2988450.2988456], https://hal.inria.fr/hal-01336912.

[50] A. C. Y. TOSSOU, C. DIMITRAKAKIS.*Algorithms for Differentially Private Multi-Armed Bandits*, in "AAAI 2016", Phoenix, Arizona, United States, February 2016, https://hal.inria.fr/hal-01234427.

[51] Z. ZHANG, B. RUBINSTEIN, C. DIMITRAKAKIS.*On the Differential Privacy of Bayesian Inference*, in "AAAI 2016", Phoenix, Arizona, United States, February 2016, https://hal.inria.fr/hal-01234215.

Conferences without Proceedings

[52] A. BÉRARD, C. SERVAN, O. PIETQUIN, L. BESACIER.*MultiVec: a Multilingual and Multilevel Representation Learning Toolkit for NLP*, in "The 10th edition of the Language Resources and Evaluation Conference (LREC)", Portoroz, Slovenia, May 2016, https://hal.archives-ouvertes.fr/hal-01335930.

[53] F. GUILLOU, R. GAUDEL, P. PREUX.*Compromis exploration-exploitation pour système de recommandation à grande échelle*, in "Conférence francophone sur l'Apprentissage Automatique (CAp'16)", Marseille, France, July 2016, https://hal.inria.fr/hal-01406439.

[54] F. STRUB, J. MARY, R. GAUDEL.*Filtrage Collaboratif Hybride avec des Auto-encodeurs*, in "Conférence francophone sur l'Apprentissage Automatique (CAp'16)", Marseille, France, July 2016, https://hal.inria.fr/hal-01406432.

Research Reports

[55] B. DANGLOT, P. PREUX, B. BAUDRY, M. MONPERRUS.*Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation*, HAL, 2016, n° hal-01378523, https://hal.archives-ouvertes.fr/hal-01378523.

Other Publications

[56] S. BUBECK, R. ELDAN, J. LEHEC.*Sampling from a log-concave distribution with Projected Langevin Monte Carlo*, January 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01428950.

[57] C. DIMITRAKAKIS, F. JARBOUI, D. PARKES, L. SEEMAN.*Multi-view Sequential Games: The Helper-Agent Problem*, December 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01408294.

[58] E. KAUFMANN.*On Bayesian index policies for sequential resource allocation*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01251606.

[59] A. R. LUEDTKE, E. KAUFMANN, A. CHAMBAZ.*Asymptotically Optimal Algorithms for Multiple Play Bandits with Partial Feedback*, June 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01338733.

[60] F. STRUB, J. MARY, R. GAUDEL.*Hybrid Collaborative Filtering with Autoencoders*, July 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01281794.

References in notes

[61] P. AUER, N. CESA-BIANCHI, P. FISCHER.*Finite-time analysis of the multi-armed bandit problem*, in "Machine Learning", 2002, vol. 47, n^{o} 2/3, p. 235–256.

[62] R. BELLMAN.*Dynamic Programming*, Princeton University Press, 1957.

[63] D. BERTSEKAS, S. SHREVE.*Stochastic Optimal Control (The Discrete Time Case)*, Academic Press, New York, 1978.

[64] D. BERTSEKAS, J. TSITSIKLIS.*Neuro-Dynamic Programming*, Athena Scientific, 1996.

[65] M. PUTERMAN.*Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley and Sons, 1994.

[66] H. ROBBINS.*Some aspects of the sequential design of experiments*, in "Bull. Amer. Math. Soc.", 1952, vol. 55, p. 527–535.

[67] R. SUTTON, A. BARTO.*Reinforcement learning: an introduction*, MIT Press, 1998.

[68] P. WERBOS.*ADP: Goals, Opportunities and Principles*, IEEE Press, 2004, p. 3–44, Handbook of learning and approximate dynamic programming.

Project-Team SPIRALS

Self-adaptation for distributed services and large software systems

IN COLLABORATION WITH: Centre de Recherche en Informatique, Signal et Automatique de Lille

IN PARTNERSHIP WITH:
Université des sciences et technologies de Lille (Lille 1)

RESEARCH CENTER
Lille - Nord Europe

THEME
Distributed Systems and middleware

Table of contents

Project-Team SPIRALS

Creation of the Team: 2014 January 01, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:
- 1.1.6. - Cloud
- 1.2.6. - Sensor networks
- 1.2.7. - Cyber-physical systems
- 1.3. - Distributed Systems
- 1.4. - Ubiquitous Systems
- 1.6. - Green Computing
- 2. - Software
- 2.1.7. - Distributed programming
- 2.5. - Software engineering
- 2.5.1. - Software Architecture & Design
- 2.5.2. - Component-based Design
- 2.5.3. - Empirical Software Engineering
- 2.5.4. - Software Maintenance & Evolution
- 2.6.2. - Middleware
- 3.1.3. - Distributed data

Other Research Topics and Application Domains:
- 4.5.1. - Green computing
- 6.1. - Software industry
- 8.1. - Smart building/home
- 8.2. - Connected city
- 8.5.2. - Crowd sourcing
- 9.4.1. - Computer science

1. Members

Research Scientist

Philippe Merle [Inria, Researcher, HDR]

Faculty Members

Lionel Seinturier [Team leader, Univ. Lille I, Professor, HDR]
Laurence Duchien [Univ. Lille I, Professor, HDR]
Martin Monperrus [Univ. Lille I, Associate Professor, HDR]
Romain Rouvoy [Univ. Lille I, Professor, HDR]
Walter Rudametkin Ivey [Univ. Lille I, Associate Professor]
Marc Sango [Univ. Lille I, ATER, until Aug 2016]

Technical Staff

Benjamin Danglot [Inria, from Sep to Nov 2016, ADT LibRepair]
Julien Duribreux [Inria, ADT CrowdLab, FUI StoreConnect]
Christophe Gourdin [Inria, from Feb 2016, PIA OCCIware]

Gérard Paligot [Inria, until Sep 2016, ADT Spoon3R]
Simon Urli [Inria, from Dec 2016, ADT LibRepair]
Antoine Veuiller [Inria, ADT CrowdLab, FUI StoreConnect]

PhD Students

Benjamin Danglot [Inria, from Dec 2016, H2020 ICT Stamp]
Yahya Al-Dhuraibi [Scalair, granted by CIFRE]
Stéphanie Challita [Inria, PIA OCCIware]
Maxime Colmant [Univ. Lille I, Lille I/ADEME]
Thomas Durieux [Inria, Microsoft Research]
María Gómez Lacruz [Inria]
Miguel Alejandro Gonzalez [CONACYT, from Oct 2016]
Sarra Habchi [Inria, from Sep 2016]
Geoffrey Hecht [Inria, Lille 1/UQAM, co-supervision with N. Moha from UQAM]
Lakhdar Meftah [Inria, from Nov 2016, co-supervision with I. Chrisment from Inria Madynes, IPL BetterNet]
Vincenzo Musco [Univ. Lille III, until Nov 2016, Lille III/Lille I, co-supervision with P. Preux from Inria Sequel]
Gustavo Sousa [Erasmus Mundus]
Amal Tahri [Orange, granted by CIFRE]
Antoine Vastel [Univ. Lille I, from Sep 2016]
Bo Zhang [Inria, PIA Datalyse]

Post-Doctoral Fellows

Fawaz Paraiso [Inria, PIA OCCIware]
Daniel Romero Acero [Inria, until Feb 2016, FP7 PaaSage]
Zhongxing Yu [Inria, from Aug 2016]

Visiting Scientists

Fernanda Madeiral Delfim [Univ. Uberlândia, PhD Student, from Dec 2016, STIC AmSud MineAPI]
Marcelo Maia [Univ. Uberlândia, Professor, Nov 2016, STIC AmSud MineAPI]
Mohamed Berkane [Univ. Constantine 2, Oct 2016]

Administrative Assistant

Aurore Hermant [Inria]

Others

Luc Delataille [Education nationale, Elève de 3ème en stage de découverte, Feb 2016]
Arnaud Deleruyelle [Inria, Bachelor Intern, from May 2016 until Aug 2016]
Dominik Duhamel [Education nationale, Elève de 3ème en stage de découverte, Jan 2016]
Tom Fleter [Education nationale, Elève de 3ème en stage de découverte, Nov 2016]
Christophe Gransart [Ifsttar, Researcher, External collaborator]
Timothée Lefebvre [Inria, Master Intern, from Jun 2016 until Jul 2016]
Maxime Opsommer [Inria, Bachelor Intern, from Apr 2016 until Jul 2016]
Romain Sommerard [Inria, Master Intern, from Mar 2016 until Aug 2016]

2. Overall Objectives

2.1. Introduction

Our research is based on two complementary fields: distributed systems and software engineering. We aim at introducing more automation in the adaptation processes of software systems, that is, transitioning from the study of adaptive systems to self-adaptive systems. In particular, we work towards two directions: self-healing software systems with data mining solutions, and self-optimizing software systems with context monitoring. These two objectives are declined for two target environments: mobile computing and cloud computing.

2.2. Scientific Foundations

Distributed software services and systems are central to many human activities, such as communication, commerce, education, defense, etc. Distributed software services consist of an ever growing number of devices, often highly heterogeneous, from cloud platforms, sensor networks, to application servers, desktop machines, and mobile devices, such as smartphones. The future of this huge number of interconnected software services has been called the Internet of Services, a vision *"where everything that is needed to use software applications is available as a service on the Internet, such as the software itself, the tools to develop the software, the platform servers, storage and communication to run the software."* [0] This pervasiveness continuously leads to new usages that in turn foster the emergence of novel requirements and concepts for new software services. Hence, it is necessary to establish new paradigms to design and execute software programs in these highly interconnected and heterogeneous environments, and it is necessary to ensure not only that these software systems can be adapted to new usages, new infrastructures, and new execution environments in the long term, but also that after the adaptation process the services still perform as expected.

This research project focuses on defining *self-adaptive* software services and middleware. From the perspective of the Internet of Services, this project fits in the vision sketched by *e.g.*, the FP8 Expert Group Services in the Future Internet [70], the NESSI Research Priorities for the next Framework Programme for Research and Technological Development FP8 [73], the Roadmap for Advanced Cloud Technologies under H2020 [71], and research roadmaps, such as [78], [69], [63].

3. Research Program

3.1. Introduction

Our research program on self-adaptive software targets two key properties that are detailed in the remainder of this section: *self-healing* and *self-optimization*.

3.2. Objective #1: Self-healing - Mining software artifacts to automatically evolve systems

Software systems are under the pressure of changes all along their lifecycle. Agile development blurs the frontier between design and execution and requires constant adaptation. The size of systems (millions of lines of code) multiplies the number of bugs by the same order of magnitude. More and more systems, such as sensor network devices, live in "surviving" mode, in the sense that they are neither rebootable nor upgradable.

Software bugs are hidden in source code and show up at development-time, testing-time or worse, once deployed in production. Except for very specific application domains where formal proofs are achievable, bugs can not be eradicated. As an order of magnitude, on 16 Dec 2011, the Eclipse bug repository contains 366,922 bug reports. Software engineers and developers work on bug fixing on a daily basis. Not all developers spend the same time on bug fixing. In large companies, this is sometimes a full-time role to manage bugs, often referred to as *Quality Assurance* (QA) software engineers. Also, not all bugs are equal, some bugs are analyzed and fixed within minutes, others may take months to be solved [75].

In terms of research, this means that: (i) one needs means to automatically adapt the design of the software system through automated refactoring and API extraction, (ii) one needs approaches to automate the process of adapting source code in order to fix certain bugs, (iii) one needs to revisit the notion of error-handling so that instead of crashing in presence of errors, software adapts itself to continue with its execution, *e.g.*, in degraded mode.

[0] http://cordis.europa.eu/fp7/ict/ssai

There is no one-size-fits-all solution for each of these points. However, we think that novel solutions can be found by using **data mining and machine learning techniques tailored for software engineering** [76]. This body of research consists of mining some knowledge about a software system by analyzing the source code, the version control systems, the execution traces, documentation and all kinds of software development and execution artifacts in general. This knowledge is then used within recommendation systems for software development, auditing tools, runtime monitors, frameworks for resilient computing, etc.

The novelty of our approach consists of using and tailoring data mining techniques for analyzing software artifacts (source code, execution traces) in order to achieve the **next level of automated adaptation** (*e.g.*, automated bug fixing). Technically, we plan to mix unsupervised statistical learning techniques (*e.g.* frequent item set mining) and supervised ones (*e.g.* training classifiers such as decision trees). This research is currently not being performed by data mining research teams since it requires a high level of domain expertise in software engineering, while software engineering researchers can use off-the-shelf data mining libraries, such as Weka [61].

We now detail the two directions that we propose to follow to achieve this objective.

3.2.1. Learning from software history how to design software and fix bugs

The first direction is about mining techniques in software repositories (*e.g.*, CVS, SVN, Git). Best practices can be extracted by data mining source code and the version control history of existing software systems. The design and code of expert developers significantly vary from the artifacts of novice developers. We will learn to differentiate those design characteristics by comparing different code bases, and by observing the semantic refactoring actions from version control history. Those design rules can then feed the test-develop-refactor constant adaptation cycle of agile development.

Fault localization of bugs reported in bug repositories. We will build a solid foundation on empirical knowledge about bugs reported in bug repository. We will perform an empirical study on a set of representative bug repositories to identify classes of bugs and patterns of bug data. For this, we will build a tool to browse and annotate bug reports. Browsing will be helped with two kinds of indexing: first, the tool will index all textual artifacts for each bug report; second it will index the semantic information that is not present by default in bug management software—*i.e.*, "contains a stacktrace"). Both indexes will be used to find particular subsets of bug reports, for instance "all bugs mentioning invariants and containing a stacktrace". Note that queries with this kind of complexity and higher are mostly not possible with the state-of-the-art of bug management software. Then, analysts will use annotation features to annotate bug reports. The main outcome of the empirical study will be the identification of classes of bugs that are appropriate for automated localization. Then, we will run machine learning algorithms to identify the latent links between the bug report content and source code features. Those algorithms would use as training data the existing traceability links between bug reports and source code modifications from version control systems. We will start by using decision trees since they produce a model that is explicit and understandable by expert developers. Depending on the results, other machine learning algorithms will be used. The resulting system will be able to locate elements in source code related to a certain bug report with a certain confidence.

Automated bug fix generation with search-based techniques. Once a location in code is identified as being the cause of the bug, we can try to automatically find a potential fix. We envision different techniques: (1) infer fixes from existing contracts and specifications that are violated; (2) infer fixes from the software behavior specified as a test suite; (3) try different fix types one-by-one from a list of identified bug fix patterns; (4) search fixes in a fix space that consists of combinations of atomic bug fixes. Techniques 1 and 2 are explored in [58] and [74]. We will focus on the latter techniques. To identify bug fix patterns and atomic bug fixes, we will perform a large-scale empirical study on software changes (also known as changesets when referring to changes across multiple files). We will develop tools to navigate, query and annotate changesets in a version control system. Then, a grounded theory will be built to master the nature of fixes. Eventually, we will decompose change sets in atomic actions using clustering on changeset actions. We will then use this body of empirical knowledge to feed search-based algorithms (*e.g.* genetic algorithms) that will look for meaningful fixes in a large fix space. To sum up, our research on automated bug fixing will try not only to point to source code locations responsible of a bug, but to search for code patterns and snippets that may constitute the skeleton of a valid patch. Ultimately, a blend of expert heuristics and learned rules will be able to produce valid source code that can be validated by developers and committed to the code base.

3.2.2. Run-time self-healing

The second proposed research direction is about inventing a self-healing capability at run-time. This is complementary to the previous objective that mainly deals with development time issues. We will achieve this in two steps. First, we want to define frameworks for resilient software systems. Those frameworks will help to maintain the execution even in the presence of bugs—*i.e.* to let the system survive. As exposed below, this may mean for example to switch to some degraded modes. Next, we want to go a step further and to define solutions for automated runtime repair, that is, not simply compensating the erroneous behavior, but also determining the correct repair actions and applying them at run-time.

Mining best effort values. A well-known principle of software engineering is the "fail-fast" principle. In a nutshell, it states that as soon as something goes wrong, software should stop the execution before entering incorrect states. This is fine when a human user is in the loop, capable of understanding the error or at least rebooting the system. However, the notion of "failure-oblivious computing" [68] shows that in certain domains, software should run in a resilient mode (*i.e.* capable of recovering from errors) and/or best-effort mode—*i.e.* a slightly imprecise computation is better than stopping. Hence, we plan to investigate data mining techniques in order to learn best-effort values from past executions (*i.e.* somehow learning what is a correct state, or the opposite what is not a completely incorrect state). This knowledge will then be used to adapt the software state and flow in order to mitigate the error consequences, the exact opposite of fail-fast for systems with long-running cycles.

Embedding search based algorithms at runtime. Harman recently described the field of search-based software engineering [62]. We think that certain search based approaches can be embedded at runtime with the goal of automatically finding solutions that avoid crashing. We will create software infrastructures that allow automatically detecting and repairing faults at run-time. The methodology for achieving this task is based on three points: (1) empirical study of runtime faults; (2) learning approaches to characterize runtime faults; (3) learning algorithms to produce valid changes to the software runtime state. An empirical study will be performed to analyze those bug reports that are associated with runtime information (*e.g.* core dumps or stacktraces). After this empirical study, we will create a system that learns on previous repairs how to produce small changes that solve standard runtime bugs (*e.g.* adding an array bound check to throw a handled domain exception rather than a spurious language exception). To achieve this task, component models will be used to (1) encapsulate the monitoring and reparation meta-programs in appropriate components and (2) support runtime code modification using scripting, reflective or bytecode generation techniques.

3.3. Objective #2: Self-optimization - Sharing runtime behaviors to continuously adapt software

Complex distributed systems have to seamlessly adapt to a wide variety of deployment targets. This is due to the fact that developers cannot anticipate all the runtime conditions under which these systems are immersed. A major challenge for these software systems is to develop their capability to continuously reason about themselves and to take appropriate decisions and actions on the optimizations they can apply to improve themselves. This challenge encompasses research contributions in different areas, from environmental monitoring to real-time symptoms diagnosis, to automated decision making. The variety of distributed systems, the number of optimization parameters, and the complexity of decisions often resign the practitioners to design monolithic and static middleware solutions. However, it is now globally acknowledged that the development of dedicated building blocks does not contribute to the adoption of sustainable solutions. This is confirmed by the scale of actual distributed systems, which can—for example—connect several thousands of devices to a set of services hosted in the Cloud. In such a context, the lack of support for smart behaviours at different levels of the systems can inevitably lead to its instability or its unavailability. In June 2012, an outage of Amazon's Elastic Compute Cloud in North Virginia has taken down Netflix, Pinterest, and

Instagram services. During hours, all these services failed to satisfy their millions of customers due to the lack of integration of a self-optimization mechanism going beyond the boundaries of Amazon.

The research contributions we envision within this area will therefore be organized as a reference model for engineering **self-optimized distributed systems** autonomously driven by *adaptive feedback control loops*, which will automatically enlarge their scope to cope with the complexity of the decisions to be taken. This solution introduces a multi-scale approach, which first privileges local and fast decisions to ensure the homeostasis [0] property of a single node, and then progressively propagates symptoms in the network in order to reason on a longer term and a larger number of nodes. Ultimately, domain experts and software developers can be automatically involved in the decision process if the system fails to find a satisfying solution. The research program for this objective will therefore focus on the study of mechanisms for **monitoring, taking decisions, and automatically reconfiguring software at runtime and at various scales**. As stated in the self-healing objective, we believe that there is no one-size-fits-all mechanism that can span all the scales of the system. We will therefore study and identify an optimal composition of various adaptation mechanisms in order to produce long-living software systems.

The novelty of this objective is to exploit the wisdom of crowds to define new middleware solutions that are able to continuously adapt software deployed in the wild. We intend to demonstrate the applicability of this approach to distributed systems that are deployed from mobile phones to cloud infrastructures. The key scientific challenges to address can be summarized as follows: *How does software behave once deployed in the wild? Is it possible to automatically infer the quality of experience, as it is perceived by users? Can the runtime optimizations be shared across a wide variety of software? How optimizations can be safely operated on large populations of software instances?*

The remainder of this section further elaborates on the opportunities that can be considered within the frame of this objective.

3.3.1. Monitoring software in the wild

Once deployed, developers are generally no longer aware of how their software behave. Even if they heavily use testbeds and benchmarks during the development phase, they mostly rely on the bugs explicitly reported by users to monitor the efficiency of their applications. However, it has been shown that contextual artifacts collected at runtime can help to understand performance leaks and optimize the resilience of software systems [77]. Monitoring and understanding the context of software at runtime therefore represent the first building block of this research challenge. Practically, we intend to investigate crowd-sensing approaches, to smartly collect and process runtime metrics (*e.g.*, request throughput, energy consumption, user context). Crowd-sensing can be seen as a specific kind of crowdsourcing activity, which refers to the capability of lifting a (large) diffuse group of participants to delegate the task of retrieving trustable data from the field. In particular, crowd-sensing covers not only *participatory sensing* to involve the user in the sensing task (*e.g.*, surveys), but also *opportunistic sensing* to exploit mobile sensors carried by the user (*e.g.*, smartphones).

While reported metrics generally enclose raw data, the monitoring layer intends to produce meaningful indicators like the *Quality of Experience* (QoE) perceived by users. This QoE reflects representative symptoms of software requiring to trigger appropriate decisions in order to improve its efficiency. To diagnose these symptoms, the system has to process a huge variety of data including runtime metrics, but also history of logs to explore the sources of the reported problems and identify opportunities for optimizations. The techniques we envision at this level encompass machine learning, principal component analysis, and fuzzy logic [67] to provide enriched information to the decision level.

3.3.2. Collaborative decision-making approaches

Beyond the symptoms analysis, decisions should be taken in order to improve the *Quality of Service* (QoS). In our opinion, collaborative approaches represent a promising solution to effectively converge towards the most appropriate optimization to apply for a given symptom. In particular, we believe that exploiting the wisdom

[0] Homeostasis is the property of a system that regulates its internal environment and tends to maintain a stable, relatively constant condition of properties [Wikipedia].

of the crowd can help the software to optimize itself by sharing its experience with other software instances exhibiting similar symptoms. The intuition here is that the body of knowledge that supports the optimization process cannot be specific to a single software instance as this would restrain the opportunities for improving the quality and the performance of applications. Rather, we think that any software instance can learn from the experience of others.

With regard to the state-of-the-art, we believe that a multi-levels decision infrastructure, inspired from distributed systems like Spotify [60], can be used to build a decentralized decision-making algorithm involving the surrounding peers before requesting a decision to be taken by more central control entity. In the context of collaborative decision-making, peer-based approaches therefore consist in quickly reaching a consensus on the decision to be adopted by a majority of software instances. Software instances can share their knowledge through a micro-economic model [56], that would weight the recommendations of experienced instances, assuming their age reflects an optimal configuration.

Beyond the peer level, the adoption of algorithms inspired from evolutionary computations, such as genetic programming, at an upper level of decision can offer an opportunity to test and compare several alternative decisions for a given symptom and to observe how does the crowd of applications evolves. By introducing some diversity within this population of applications, some instances will not only provide a satisfying QoS, but will also become naturally resilient to unforeseen situations.

3.3.3. Smart reconfigurations in the large

Any decision taken by the crowd requires to propagate back to and then operated by the software instances. While simplest decisions tend to impact software instances located on a single host (*e.g.*, laptop, smartphone), this process can also exhibit more complex reconfiguration scenarios that require the orchestration of various actions that have to be safely coordinated across a large number of hosts. While it is generally acknowledged that centralized approaches raise scalability issues, we think that self-optimization should investigate different reconfiguration strategies to propagate and apply the appropriate actions. The investigation of such strategies can be addressed in two steps: the consideration of *scalable data propagation protocols* and the identification of *smart reconfiguration mechanisms*.

With regard to the challenge of scalable data propagation protocols, we think that research opportunities encompass not only the exploitation of gossip-based protocols [59], but also the adoption of publish/subscribe abstractions [64] in order to decouple the decision process from the reconfiguration. The fundamental issue here is the definition of a communication substrate that can accommodate the propagation of decisions with relaxed properties, inspired by *Delay Tolerant Networks* (DTN), in order to reach weakly connected software instances. We believe that the adoption of asynchronous communication protocols can provide the sustainable foundations for addressing various execution environments including harsh environments, such as developing countries, which suffer from a partial connectivity to the network. Additionally, we are interested in developing the principle of *social networks of applications* in order to seamlessly group and organize software instances according to their similarities and acquaintances. The underlying idea is that grouping application instances can contribute to the identification of optimization profiles not only contributing to the monitoring layer, but also interested in similar reconfigurations. Social networks of applications can contribute to the anticipation of reconfigurations by exploiting the symptoms of similar applications to improve the performance of others before that problems actually happen.

With regard to the challenge of smart reconfiguration mechanisms, we are interested in building on our established experience of adaptive middleware [72] in order to investigate novel approaches to efficient application reconfigurations. In particular, we are interested in adopting seamless micro-updates and micro-reboot techniques to provide in-situ reconfiguration of pieces of software. Additionally, the provision of safe and secured reconfiguration mechanisms is clearly a key issue that requires to be carefully addressed in order to avoid malicious exploitation of dynamic reconfiguration mechanisms against the software itself. In this area, although some reconfiguration mechanisms integrate transaction models [65], most of them are restricted to local reconfigurations, without providing any support for executing distributed reconfiguration transactions.

Additionally, none of the approached published in the literature include security mechanisms to preserve from unauthorized or malicious reconfigurations.

4. Application Domains

4.1. Introduction

Although our research is general enough to be applied to many application domains, we currently focus on applications and distributed services for the retail industry and for the digital home. These two application domains are supported by a strong expertise in mobile computing and in cloud computing that are the two main target environments on which our research prototypes are build, for which we are recognized, and for which we have already established strong collaborations with the industrial ecosystem.

4.2. Distributed software services for the retail industry

This application domain is developed in relation with the PICOM (*Pôle de compétitivité Industries du Commerce*) cluster. We have established strong collaborations with local companies in the context of former funded projects, such as Cappucino and Macchiato, which focused on the development of a new generation of mobile computing platforms for e-commerce. We are also involved in the Datalyse and OCCIware funded projects that define cloud computing environments with applications for the retail industry. Finally, our activities in terms of crowd-sensing and data gathering on mobile devices with the APISENSE® platform share also applications for the retail industry.

4.3. Distributed software services for the digital home

We are developing new middleware solutions for the digital home, in particular through our long standing collaboration with Orange Labs. We are especially interested in developing energy management and saving solutions with the POWERAPI software library for distributed environments such the ones that equip digital homes. We are also working to bridge the gap between distributed services hosted on home gateways and distributed services hosted on the cloud to be able to smoothly transition between both environments. This work is especially conducted with the SALOON platform.

5. Highlights of the Year

5.1. Highlights of the Year

Makitoo, the start-up company founded by Martin Monperrus and Nicolas Petitprez received the Bpifrance *Création d'entreprise innovante* award, which is a major award in France for startup companies, in the category *Création-développement*.

Makitoo won also a NETVA award from the French ministry of foreign affairs in order to develop its activities in the USA.

Romain Rouvoy has been awarded a Institut Universitaire de France (IUF) junior fellowship for 5 years (2016-21). IUF is an excellence award that is only granted to the top 2% of faculty members in French universities. The award recognizes the excellence of the research activities conducted by Romain Rouvoy.

Laurence Duchien has been elected for a 2-year term in the executive committee of the IEEE Technical Council on Software Engineering (TCSE). The IEEE TCSE helps advance software engineering research and practice. The executive committee determines TCSE policy and the nature of TCSE activities.

6. New Software and Platforms

6.1. APISENSE

KEYWORDS: Mobile sensing - Crowd-sensing - Crowd-sourcing - Android
FUNCTIONAL DESCRIPTION

APISENSE platform is a software solution to collect various contextual information from Android devices (client application) and automatically upload collected data to a server (deployed as a SaaS). APISENSE is based on a Cloud computing infrastructure to facilitate datasets collection from significant populations of mobile users for research purposes.

- Participants: Nicolas Haderer, Romain Rouvoy, Christophe Ribeiro, Julien Duribreux and Antoine Veuiller
- Partner: Université Lille 1
- Contact: Romain Rouvoy
- URL: http://www.apisense.io

6.2. Nopol

KEYWORD: Automatic software repair
FUNCTIONAL DESCRIPTION

Nopol is an automatic software repair tool for buggy conditional statements (i.e., if-then-else statements) in Java programs. Nopol takes a buggy program as well as a test suite as input and generates a patch with a conditional expression as output. The test suite is required to contain passing test cases to model the expected behavior of the program and at least one failing test case that reveals the bug to be repaired. The process of Nopol consists of three major phases. First, Nopol employs angelic fix localization to identify expected values of a condition during the test execution. Second, runtime trace collection is used to collect variables and their actual values, including primitive data types and objected-oriented features (e.g., nullness checks), to serve as building blocks for patch generation. Third, Nopol encodes these collected data into an instance of a Satisfiability Modulo Theory (SMT) problem, then a feasible solution to the SMT instance is translated back into a code patch.

- Contact: Martin Monperrus
- URL: https://github.com/SpoonLabs/nopol/

6.3. PowerAPI

KEYWORD: Energy management
FUNCTIONAL DESCRIPTION

PowerAPI is a library for monitoring the energy consumption of software systems.

PowerAPI differs from existing energy process-level monitoring tool in its software orientation, with a fully customizable and modular solution that let the user to precisely define what he/she wants to monitor. PowerAPI is based on a modular and asynchronous event-driven architecture using the Akka library. PowerAPI offers an API which can be used to define requests about energy spent by a process, following its hardware resource utilization (in term of CPU, memory, disk, network, etc.).

- Participants: Romain Rouvoy, Adel Noureddine, Loic Huertas and Maxime Colmant
- Contact: Romain Rouvoy
- URL: http://www.powerapi.org

6.4. SPOON

KEYWORDS: Java - Code analysis
FUNCTIONAL DESCRIPTION

Spoon is an open-source library that enables you to transform (see below) and analyze Java source code (see example) . Spoon provides a complete and fine-grained Java metamodel where any program element (classes, methods, fields, statements, expressions...) can be accessed both for reading and modification. Spoon takes as input source code and produces transformed source code ready to be compiled.

- Participants: Nicolas Petitprez, Martin Monperrus, Lionel Seinturier and Gérard Paligot
- Contact: Martin Monperrus
- URL: http://spoon.gforge.inria.fr

6.5. Saloon

KEYWORDS: Feature Model - Software Product Line - Cloud computing - Model-driven engineering - Ontologies
FUNCTIONAL DESCRIPTION

Saloon is a framework for the selection and configuration of Cloud providers according to application requirements. The framework enables the specification of such requirements by defining ontologies. Each ontology provides a unified vision of provider offers in terms of frameworks, databases, languages, application servers and computational resources (i.e., memory, storage and CPU frequency). Furthermore, each provider is related to a Feature Model (FM) with attributes and cardinalities, which captures its capabilities. By combining the ontology and FMs, the framework is able to match application requirements with provider capabilities and select a suitable one. Specific scripts to the selected provider are generated in order to enable its configuration.

- Participants: Clement Quinton, Daniel Romero Acero, Laurence Duchien, Lionel Seinturier and Romain Rouvoy
- Partner: Université Lille 1
- Contact: Lionel Seinturier
- URL: https://gitlab.irisa.fr/drome00A/saloon

7. New Results

7.1. Change Impact Analysis

In [21], we have proposed a novel evaluation technique for change impact analysis (CIA). CIA is a prediction problem that, given a source code element in a program, determines the other source code elements impacted if one changes this original source code element. Given the large size of the element space in complex programs, this prediction requires a trade-off between different dimensions: precision, completeness, time. The novelty of the result lies in the use of mutation analysis to study simultaneously these three dimensions. This result is backed by an empirical evaluation performed on 10 open-source Java programs and 5 mutation operators, which enabled to generate 17,000 mutants and study how the error they introduce propagates. This result has been achieved in the context of the PhD thesis, defended in November 2016, of Vicenzo Musco [15].

7.2. Learning Power Models for Distributed and Virtualized Environments

Energy efficiency is a major concern for modern ICT infrastructures. The a priori estimation of the level of energy consumed by a given service is a difficult problem given the intricate nature of hardware and software that are involved. Consequently, even before considering saving, measuring the exact amount of energy consumed by a given software service or process is required. Over the last few years, a dozen of ad hoc power models have been proposed in the literature. Nevertheless they cannot cope with the constant evolution of software and hardware architecture. We have therefore defined and implemented a toolkit that automatically learns the power models of a given architecture, independently of the features and the complexity it exhibits. This toolkit considers traditional distributed environment as well as virtualized, cloud-based ones. This result has been achieved in the context of the PhD thesis, defended in November 2016, of Maxime Colmant [11].

7.3. Crowdmining to Increase the Quality of Software Systems

Modern software systems, especially in the open source world, are more and more part of ecosystems where large quantities of data about these systems are available. These data may come for example from application stores (e.g. Google Play Store or Apple Store for mobile applications), forges (e.g. GitHub), or from the usage conditions experienced by users of these software systems. This large amount of data enables to unlock some specific challenges where knowledge about the software systems can be automatically mined and learnt. In this domain, we obtained new results on the mining of mobile software antipatterns on a crowd of mobile applications and their versions to study their impact on resource consumption [32]. This result has been achieved in the context of the PhD thesis, defended in November 2016, of Geoffrey Hecht [13]. We also consider the crowd of mobile devices and users to detect and reproduce application crashes in the wild. By leveraging our results in the domain of in-breath monitoring, we use the APISENSE® platform (see Section 6.1) to collect extended crash reports that can be aggregated to infer the minimal execution path that lead to a crash [28]. This result has been achieved in the context of the PhD thesis, defended in December 2016, of María Gomez Lacruz [12]. These results are also in relation with our activities in the context of the SOMCA associated team (see Section 9.4).

7.4. Self-Optimization of Virtualized Environments

Elasticity is a major property of virtualized computing environments. In this domain, we especially work at the infrastructure and platform levels of a cloud computing system where we obtained two results that enable to better self-optimize the consumed resources. At the infrastructure level, we proposed CloudGC, a new middleware service for suspending, resuming, and recycling idle virtual machines. The algorithm has been implemented on top of the OpenStack cloud operating system. At the platform level, we proposed a new self-balancing approach to dynamically optimize the performance of the Hadoop framework for the distributed storage and processing of large data sets. These results have been achieved in the context of the PhD thesis, defended in December 2016, of Bo Zhang [16].

8. Bilateral Contracts and Grants with Industry

8.1. ip-label

Participant: Romain Rouvoy [correspondant].

A software exploitation license of the APISENSE® crowd-sensing platform has been sold to the ip-label company. They use this platform as a solution to monitor the quality of the GSM signal in the wild. The objective is to provide developers and stakeholders with a feedback on the quality of experience of GSM connection depending on their location.

8.2. Orange Labs

Participants: Laurence Duchien [correspondant], Amal Tahri.

This collaboration aims at bridging the gap between home networks and cloud environments for the design, the provisioning and the administration of distributed services. The purpose is to define solutions, essentially software design tools and runtime infrastructures, for the seamless migration of distributed applications and services between home networks and cloud environments. The envisioned approach is based on the research activities that we are conducting in the domain of software product lines.

This collaboration is conducted in the context of the ongoing PhD thesis of Amal Tahri.

8.3. Scalair

Participants: Yahya Al-Dhuraibi, Philippe Merle [correspondant].

This collaboration aims at proposing a framework to deal with elasticity in cloud computing environments. This framework must cover all kind of resources, IaaS, PaaS, SaaS, must provide a solution for interoperability between different clouds and virtualization technologies, and must enable the specification and composition of reactive and predictive strategies.

This collaboration is conducted in the context of the ongoing PhD thesis of Yahya Al-Dhuraibi.

8.4. OpenIO

Participants: Philippe Merle, Romain Rouvoy [correspondant], Lionel Seinturier.

This collaboration aims at producing a scientific and technical state-of-the-art analysis of solutions for the large scale storage of object data in the cloud. This study aims at identifying the main properties of the existing solutions, and their differentiating factors. The solution provided by the OpenIO company will be positioned with respect to the other solutions existing on the market and in the international scientific community. Starting from this state-of-the-art, several perspectives will be identified and a research roadmap will be defined.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Région Nord-Pas De Calais

9.1.1.1. Citizen Awareness and Contribution to Air Quality Monitoring

Participants: Daniel Romero Acero, Romain Rouvoy [correspondant], Lionel Seinturier.

This is a 3-year project in the context of the so-called "Chercheur citoyen" program that started in 2015. The partners are LISIC/Université Côte d’Opale (leader), ATMO Nord-Pas De Calais, Association Bâtisseurs d’Economie Solidaire. This project targets the distributed monitoring of air quality with crowd-sensing solutions obtained via sensors connected to smart devices. We aim at inciting citizens to perform their own measures, and to obtain thanks to GPS geo-localisation a large-scale database and a dynamic fine-grained cartography of air quality. This project takes advantage of the APISENSE® crowdsensing platform (see Section 6.1).

9.1.2. Inria Lille - Nord Europe

9.1.2.1. ADT Spoon3R

Participants: Gérard Paligot, Martin Monperrus [correspondant].

ADT Spoon3R (2014–16) is a technology development initiative supported by the Inria Lille - Nord Europe Center that aims at supporting the development of the SPOON software library. (see Section 6.4) Spoon3R aims at extending SPOON with the features defined in the context of our research activities on automated software repair.

9.1.2.2. ADT LibRepair

Participants: Benjamin Danglot, Martin Monperrus [correspondant], Simon Urli.

ADT LibRepair (2016–18) is a technology development initiative supported by the Inria Lille - Nord Europe Center that aims at supporting the development of an integrated library of automated software repair algorithms and techniques. This ADT builds on our results about with the Astor, Nopol and NpeFix that have been obtained in the context of the defended PhD theses of Matias Martinez [66] and Benoit Cornu [57].

9.1.2.3. North European Lab LLEX

Participants: Martin Monperrus [correspondant], Lionel Seinturier.

North European Lab LLEX (2015–17) is an international initiative supported by the Inria Lille - Nord Europe Center that takes place in the context of a collaboration between Inria and University College London. LLEX deals with research on automatic diagnosis and repair of software bugs. Automatic software repair is the process of fixing software bugs automatically An automatic software repair system fixes software bugs with no human intervention. The goal of automatic software repair is to save maintenance costs and to enable systems to be more resilient to bugs and unexpected situations. This research may dramatically improve the quality of software systems. The objective of the partnership is to work on the automated diagnosis of exceptions with a focus on null pointer exceptions.

9.1.2.4. LEDA

Participant: Philippe Merle [correspondant].

LEDA (2013–16) Laboratoire d'Expérimentation et de Démonstrations Ambiantes is a demonstration space allocated by the Inria Lille - Nord Europe Center whose goal is to show the scientific results of the Spirals team in the domains of distributed systems, adaptable middleware, software product lines, green computing, and ambiant computing. These results are illustrated around the scenario of a mock digital home.

9.2. National Initiatives

9.2.1. ANR

9.2.1.1. ANR BottleNet

Participants: Romain Rouvoy [correspondant], Walter Rudametkin Ivey, Lionel Seinturier.

BottleNet is a 48-month project funded by ANR that started on October 2015. The objective of BottleNet is to deliver methods, algorithms, and software systems to measure Internet Quality of Experience (QoE) and diagnose the root cause of poor Internet QoE. Our goal calls for tools that run directly at users' devices. We plan to collect network and application performance metrics directly at users' devices and correlate it with user perception to model Internet QoE, and to correlate measurements across users and devices to diagnose poor Internet QoE. This data-driven approach is essential to address the challenging problem of modeling user perception and of diagnosing sources of bottlenecks in complex Internet services. BottleNet will lead to new solutions to assist users, network and service operators as well as regulators in understanding Internet QoE and the sources of performance bottleneck.

9.2.1.2. ANR SATAS

Participants: Philippe Merle [correspondant], Romain Rouvoy, Lionel Seinturier.

SATAS is a 48-month project funded by ANR that started on October 2015. SATAS aims to advance the state of the art in massively parallel SAT solving with a particular eye to the applications driving progress in the field. The final goal of the project is to be able to provide a "pay as you go" interface to SAT solving services, with a particular focus on its power consumption. This project will extend the reach of SAT solving technologies, daily used in many critical and industrial applications, to new application areas, which were previously considered too hard, and lower the cost of deploying massively parallel SAT solvers on the cloud.

9.2.2. Competitivity Clusters

9.2.2.1. FUI StoreConnect

Participants: Julien Duribreux, Romain Rouvoy, Lionel Seinturier [correspondant], Antoine Veuiller.

StoreConnect is a 24-month project funded by FUI and labelized by the PICOM (Pôle des Industries du COMmerce) competitivity cluster which has started in September 2016. The partners are Neosensys (leader), Tevolys, Ubudu, Smile, STIME, Leroy Merlin, Insiteo, Inria Spirals, Inria Fun, Inria Stars. The goal of the project is to define a modular multi-sensors middleware platform for indoor geolocation.

9.2.3. Programme Investissement d'Avenir (PIA)

9.2.3.1. PIA Datalyse

Participants: Romain Rouvoy, Lionel Seinturier [correspondant], Bo Zhang.

Datalyse is a 42-month project of the Programme Investissement d'Avenir Cloud Computing 3rd call for projects. The project started in May 2013. The partners are Eolas (leader), Business & Decision, Groupement des Mousquetaires, Université Grenoble 1, Université Lille 1, Inria, Université Montpellier 2. The project aims at defining an elastic cloud computing infrastructure for processing big volumes of data. The originality of the project is to consider jointly data generated by users and by the infrastructure, and to correlate data at these two levels.

9.2.3.2. PIA OCCIware

Participants: Romain Rouvoy, Philippe Merle [correspondant], Lionel Seinturier.

OCCIware is a 36-month project of the Programme Investissement d'Avenir Cloud Computing and Big Data 4th call for projects. The project started in December 2014. The partners are Open Wide (leader), ActiveEon SA, CSRT, Institut Mines-Télécom/Télécom SudParis, Inria, Linagora GSO, Obeo, OW2 Consortium, Pôle Numérique, and Université Joseph Fourier - Grenoble. The project aims at defining a formal framework for managing every digital resources in the clouds, based on *Open Cloud Computing Interface* (OCCI) recommendations from *Open Grid Forum* (OGF).

9.2.4. Inria National Initiatives

9.2.4.1. Inria ADT Focus CrowdLab

Participants: Julien Duribreux, María Gómez Lacruz, Romain Rouvoy [correspondant], Antoine Veuiller.

The purpose of the ADT Focus CrowdLab (2014–2016) is to strengthen the technological part of the Metroscope consortium and to promote the APISENSE® crowd-sensing platform (see Section 6.1) as a reference platform fo gathering mobile data within the scientific community. The CrowdLab project focuses on three stringent goals: (1) consolidating the current technological solutions, (2) technical and logistical support of the research activities initiated in different scientific domains, and (3) the improvement of security and anonymity of collected data. In addition to the Metroscope consortium, the Inria research teams participating of the ADT Focus CrowdLab project are: Spirals (coordinator), Madynes, Diana, Muse.

9.2.4.2. Inria IPL BetterNet

Participants: Lakhdar Meftah, Romain Rouvoy [correspondant].

BetterNet (2016–19) aims at building and delivering a scientific and technical collaborative observatory to measure and improve the Internet service access as perceived by users. In this Inria Project Lab, we will propose new original user-centered measurement methods, which will associate social sciences to better understand Internet usage and the quality of services and networks. Our observatory can be defined as a vantage point, where: (1) tools, models and algorithms/heuristics will be provided to collect data, (2) acquired data will be analyzed, and shared appropriately with scientists, stakeholders and civil society, (3) and new value-added services will be proposed to end-users. IPL BetterNet is lead by Isabelle Chrisment (Inria Madynes), with the participation of the Diana, Dionysos, Inria Chile, Muse, and Spirals Inria project-teams.

9.2.5. Others

9.2.5.1. CNRS INS2I JCJC FPDefendor

Participant: Walter Rudametkin Ivey [correspondant].

FPDefendor is a 12-month project funded by the CNRS INS2I institute. The JCJC program targets young researchers. Walter Rudametkin is the recipient of such a grant. The project aims at better understanding browser fingerprinting, its risks to privacy, and to provide measures to detect it and effective countermeasures to mitigate it. The proposal brings together software engineering, security and privacy, and formal verification to propose a platform that uses dynamic reconfiguration as a means to evade fingerprint tracking.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

Program: FP7 ICT.

Project acronym: PaaSage.

Project title: Model Based Cloud Platform Upperware.

Duration: October 2012–September 2016.

Coordinator: ERCIM.

Other partners: ERCIM (Fr), SINTEF (No), STFC (UK), U. of Stuttgart (De), Inria (Fr), CETIC (Be), FORTH (El), Be.Wan (Be), EVRY Solutions (No), SysFera (Fr), Flexiant (UK), Lufthansa Systems AG (De), Gesellschaft fur wissenschaftliche Datenverarbeitung mbh Gottingen (De), Automotive Simulation Center Stuttgart (De).

Abstract: Cloud computing is a popular and over-hyped concept in ICT. The concept of infinitely scalable elastic resources changing without complex systems administration and paying only for resources used is attractive. These benefits are not immediately realizable. Within organisation benefits are realizable at considerable cost. IaaS (*Infrastructure-as-a-Service*) public Clouds have different interfaces and conditions of use thus for an organisation to "scale out" requires considerable investment using skilled technical staff. The business need is to allow organisations to "scale out" from their private Cloud to public Clouds without a technical chasm between. This cannot easily be achieved. Aligned with the EU strategic direction of an open market for services, SOA (*Service-Oriented architecture*) offers a way to virtualize across heterogeneous public Clouds and organizational private Clouds. It opens a market for European SMEs to provide services to be utilized (and paid for) by business applications and for all organisations to benefit from a catalogue of services that can be used across the environment. PaaSage will deliver an open and integrated platform, to support both deployment and design of Cloud applications, together with an accompanying methodology that allows model-based development, configuration, optimisation, and deployment of existing and new applications independently of the existing underlying Cloud infrastructures. Specifically it will deliver an IDE (*Integrated Development Environment*) incorporating modules for design time and execution time optimisation of applications specified in the Cloud Modeling Language (Cloud ML), execution-level mappers and interfaces and a metadata database.

Participants: Laurence Duchien, Daniel Romero Acero, Romain Rouvoy, Lionel Seinturier [correspondant].

Program: FP7 FET.

Project acronym: DIVERSIFY.

Project title: More software diversity. More adaptivity in CAS.

Duration: 36 months (2013–16).

Coordinator: Inria.

Other partners: SINTEF (Norway), Trinity College Dublin (Ireland), University of Rennes 1 (France).

Abstract: DIVERSIFY explores diversity as the foundation for a novel software design principle and increased adaptive capacities in CASs (*Collective Adaptive Systems*). Higher levels of diversity in the system provide a pool of software solutions that can eventually be used to adapt to unforeseen situations at design time. The scientific development of DIVERSIFY is based on a strong analogy with ecological systems, biodiversity, and evolutionary ecology. DIVERSIFY brings together researchers from the domains of software-intensive distributed systems and ecology in order to translate ecological concepts and processes into software design principles.

Participants: Martin Monperrus [correspondant].

Program: H2020 ICT-10-2016.

Project acronym: STAMP.

Project title: Software Testing Amplification.

Duration: 36 months (2016–19).

Coordinator: Inria.

Other partners: ActiveEon (France), Atos (Spain), Engineering (Italy), OW2 (France), SINTEF (Norway), TellU (Norway), TU Delft (The Netherlands), XWiki (France).

Abstract: By leveraging advanced research in automatic test generation, STAMP aims at pushing automation in DevOps one step further through innovative methods of test amplification. It will reuse existing assets (test cases, API descriptions, dependency models), in order to generate more test cases and test configurations each time the application is updated. Acting at all steps of development cycle, STAMP techniques aim at reducing the number and cost of regression bugs at unit level, configuration level and production stage.

Participants: Benjamin Danglot, Martin Monperrus [correspondant].

Program: H2020 JU Shift2Rail.

Project acronym: X2Rail-1.

Project title: Start-up activities for Advanced Signalling and Automation System.

Duration: 36 months (2016–19).

Coordinator: Siemens.

Other partners: 19 partners, among others Bombardier, Siemens, Thales, IRT Railenium.

Abstract: Our contribution to the project is focused on adaptive communication middleware for cyber-physical railway systems.

Participants: Lionel Seinturier [correspondant].

9.3.2. Collaborations in European Programs, Except FP7 & H2020

Program: EUREKA Celtic-Plus.

Project acronym: SENDATE.

Project title: SEcure Networking for a DATa Center Cloud in Europe.

Duration: 36 months (2016–19).

Coordinator: Nokia.

Other partners: 50+ partners in Finland, France, Germany, Norway, and Sweden. Selected partners involved: Nokia, Orange.

Abstract: The project addresses the convergence of telecommunication networks and IT in the context of distributed data centers. We are involved in the TANDEM subproject that targets the infrastructure of such a distributed system. More specifically, we are studying new approaches in terms of software engineering and component-based solutions for enabling this convergence of network and IT.

Participants: Lionel Seinturier [correspondant].

9.4. International Initiatives

9.4.1. Inria Associate Teams Not Involved in an Inria International Labs

9.4.1.1. SOMCA

Title: Self-Optimization of Service Oriented Architectures for Mobile and Cloud Applications

International Partner (Institution - Laboratory - Researcher):

Université du Québec À Montréal (Canada) - LATECE - Naouel MOHA

Start year: 2014

See also: http://sofa.uqam.ca/somca.php

The long-term goal of this research program is to propose a novel and innovative methodology embodied in an software platform, to support the runtime detection and correction of anti-patterns in large-scale service-oriented distributed systems in order to continuously optimize their quality of service. One originality of this program lies in the dynamic nature of the service-oriented environments and the application on emerging frameworks for embedded and distributed systems (e.g., Android/iOS for mobile devices, PaaS/SaaS for Cloud environments), and in particular mobile systems interacting with remote services hosted on the Cloud.

9.4.2. Participation in Other International Programs

9.4.2.1. STIC AmSud - Project MineAPI

Participants: María Gómez Lacruz, Martin Monperrus [correspondant], Vincenzo Musco, Gérard Paligot, Romain Rouvoy.

MineAPI is a STIC AmSud project (2015–16) between with University Diego Portales, Santiago, Chile, and Federal University of Uberlândia, Brazil. The coordinator on the French side is Damien Cassou from Inria Rmod. The project aims at facilitating the usage of frameworks and application programming interfaces (APIs) by mining software repositories. Our intuition is that mining reveals how existing projects instantiate these frameworks. By locating concrete framework instantiations in existing projects, we can recommend to developers the concrete procedures for how to use a particular framework for a particular task in a new system. Our project also tackles the challenge of adapting existing systems to new versions of a framework or API by seeking repositories for how other systems adapted to such changes.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Prof. Marcelo Maia, from the Federal University of Uberlândia, Brazil, visited us for 1 week in November 2016 in the context of the MineAPI project.

Fernanda Madeiral Delfim, PhD Student from the Federal University of Uberlândia, Brazil, started a 6-month visit in December 2016 in the context of the MineAPI project.

Mohamed Berkane, associate professor at the University Constantine 2, Algeria, visited us for 1 month in October 2016.

9.5.1.1. Research Stays Abroad

María Gómez spent 4 months from January to April 2016 at Universität Hamburg in the research group of Prof. Walid Maalej.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Laurence Duchien has been General Chair of (EJCP'2016), that is an international summer school in the domain of software engineering. The summer school was held in Lille from 27 June to 1 July 2016.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Laurence Duchien

International Systems and Software Product Line Conference, Vision Track (SPLC)

IEEE Conference on Software Engineering Education & Training (CSEE&T)

International Workshop on Software Engineering for Systems-of-Systems (SESoS)

European Conference on Software Architecture (ECSA)

Philippe Merle

International Symposium on Security in Computing and Communications (SSCC)

International Workshop on Adaptive and Reflective Middleware (ARM)

Informatique des Organisation et Systèmes d'Information et de Décision (INFORSID)

Workshop on CrossCloud Infrastructures & Platforms (CrossCloud)

International Conference on Cloud Computing and Services Science (CLOSER), Special Session on Experiences with OCCI

Martin Monperrus

International Symposium on Software Testing and Analysis (ISSTA)

International Conference on Software Maintenance and Evolution (ICSME)

Romain Rouvoy

ACM/IFIP/USENIX Middleware conference (Middleware)

ACM Symposium on Applied Computing (SAC), track on Dependable, Adaptive, and Trustworthy Distributed Systems (DADS)

International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC)

International Conference on Ambient Systems, Networks and Technologies (ANT)

International Conference on Formal Aspects of Component Software (FACS)

IEEE International Workshop on Big Data Management for the Internet of Things (BIOT)

International Workshop on Meta-Programming Techniques and Reflection (Meta)

Lionel Seinturier

International Conference on Service Oriented Computing (ICSOC)

ACM/IEEE Joint Conference WICSA and CompArch (ICSA)

ACM Symposium on Applied Computing (SAC), tracks Software Architecture Theory Technology and Applications (SA-TTA), Operating Systems (OS)

IEEE International Symposium on Software Crowdsourcing (ISSC)

Euromicro Conference on Software Engineering and Advanced Applications (SEAA), track Model-based development Components and Services (MOCS)

International Conference on Software Paradigm Trends (ICSOFT-PT)

International Cloud Forward Conference (CF)

International Workshop on Interplay of Security, Safety and System/Software Architecture (ISSA)

International Workshop on Models@run.time @ MODELS

International Workshop on Models@run.time @ ICAC

Workshop on Software Engineering for Sustainable Systems (SE4SuSy)

International Workshop on Advanced Information Systems for Enterprises (IWAISE)

Workshop on Model-driven Engineering for Component-based Software Systems (ModComp)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Martin Monperrus is member of the editorial board of the international journal Springer Empirical Software Engineering (IF-2015: 1.933).

Romain Rouvoy is member of the editorial board of the journal Lavoisier Technique et Science Informatiques (TSI).

Lionel Seinturier is editor for software engineering of the ISTE-Wiley Computer Science and Information Technology book collection.

10.1.3.2. Reviewer - Reviewing Activities

Laurence Duchien: Lavoisier Technique et Science Informatiques (TSI), Wiley Software Practice and Experience (SPE), Elsevier Journal of Software and Systems (JSS).

Martin Monperrus: Springer Empirical Software Engineering (EMSE), Elsevier Information and Software Technology (IST), IEEE Transactions on Software Engineering (TSE).

Romain Rouvoy: Wiley Journal of Software: Evolution and Process (JSME), Elsevier Journal of Pervasive and Mobile Computing (PMC), Elsevier Journal of Computer Networks (COMNET), Elsevier Journal of Computer Communications (COMCOM), Elsevier Journal of Network and Computer Applications (JNCA), Elsevier Journal on Science of Computer Programming (SCICO).

Lionel Seinturier: ACM Computing Surveys, Elsevier Information and Software Technology (IST), Wiley Software Practice and Experience (SPE), Elsevier Science of Computer Programming (SCP), Springer International Journal on Software and Systems Modeling (SoSyM), Springer Knowledge and Information Systems (KAIS), Elsevier Computers & Electrical Engineering (CAEE), Oxford The Computer Journal, Revue Africaine de la Recherche en Informatique et Mathématiques Appliquées (ARIMA), Inderscience International Journal of Computer Applications in Technology (IJCAT).

10.1.4. Invited Talks

Romain Rouvoy gave on keynote at ECAAS 2016 in the 1st workshop on Engineering Context-Aware Applications and Services, on mobile crowdsensing.

Walter Rudametkin gave an invited talk on browser fingerprinting at CiComp 2016, the 8th International Congress on Computational Sciences.

Lionel Seinturier gave an webinar at the West University of Timisoara, Romania, on High-level Language Support for Reconfiguration Control in Component-Based Architectures.

10.1.5. Leadership within the Scientific Community

Martin Monperrus is the co-head of the "Groupe de Travail Génie Logiciel Empirique" of the GDR GPL.

Romain Rouvoy is the co-head of the "Groupe de Travail Génie Logiciel pour les Systèmes Cyber-physiques" of the GDR GPL.

10.1.6. Scientific Expertise

Laurence Duchien was member of the recruitment committee for Chaire Inria at Ecole des Mines de Nantes, and deputy head of the recruitment committee for Chargés de Recherche at Inria Lille - Nord Europe. She did some scientific expertises for Direction des Relations Internationales of the French Ministry of Research.

Martin Monperrus has evaluated grants for the Natural Sciences and Engineering Research Council of Canada (NSERC).

Romain Rouvoy has evaluated Discovery Grants for the Natural Sciences and Engineering Research Council of Canada (NSERC) and research proposals for the STIC-AmSud Program.

Lionel Seinturier was member of the recruitment committees for professors in computer science at the University of Bordeaux, and the University of Grenoble. He was scientific expert for ANRT and University Paris 6. He was member of the Agence Nationale de la Recherche (ANR) Scientific Evaluation Committee for Software and Network (CES25).

10.1.7. Research Administration

Laurence Duchien is member of the Scientific Advisory Board for the Certus Center of the Simula Lab Norway (2014-2016). She is member of the CNRS CoCNRS section 6 committee, and of the "bureau" of this committee.

Philippe Merle is member of the scientific board of Inria. He is in charge of creating the Lille's instance of the Irill research and innovation institute on free software. He is president of the CUMI (Comité des Utilisateurs des Moyens Informatiques), alternate member of the Comité de centre, member of the CLHSCT (Comité Local d'Hygiène, de Sécurité et de Conditions de Travail).

Romain Rouvoy is member of the CLDD (Commission Locale de Développement Durable), member of the CUB (Commission des Utilisateurs du Bâtiment), member of the CER (Commission Emplois Recherche), and member of the Horizon Startup committee for the Inria Lille - Nord Europe research center.

Lionel Seinturier is president of the CDT (Commission Développement Technologique), member of the BCP (Bureau du Comité des Projets), and scientific correspondant for DPEI (Direction Partenariat Europe et International), for the Inria Lille - Nord Europe research center. He heads the committee (so-called "vivier 27 rang A") that selects members of recruitment committees in Computer Science at the University of Lille.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Laurence Duchien teaches at the University of Lille 1, IEEA faculty. She heads the Carrières et Emplois service and is referent for the professional insertion in the PhD program in Computer Science at ComUE University Lille Nord de France. She is Director of Doctoral Studies for Computer Science in the Doctoral School Engineering Science (SPI) - ComUE Lille Nord de France.

Software Project Management, 50h, Level M2, Master MIAGE

Design of distributed applications, 42h, Level M1, Master of Computer Science

Software Product Lines, 8h, Level M2, Master of Computer Science

Research and Innovation Initiation, 22h, Level M2 IAGL, Master of Computer Science

Tutoring Internship, 16h, Level M2, Master of Computer Science

Martin Monperrus teaches at the University of Lille 1, IEEA faculty. He heads the IAGL specialty of the Master of Computer Science at the University of Lille 1.

Introduction to programming, 48h, Level L1, Licence of Computer Science

Object-oriented design, 39h, Level L3, Licence of Computer Science

Automated software engineering, 40h, Level M2 IAGL, Master of Computer Science

Romain Rouvoy teaches at the University of Lille 1, IEEA faculty. He heads the Master of Computer Science program at the University of Lille 1. He supervises the Agil-IT Junior Enterprise.

Infrastructures et Frameworks Internet, 32h, Level M2 IAGL, Master of Computer Science

Innovation & Initiation à la Recherche, 14h, Level M2 IAGL, Master of Computer Science speciality

Outils pour la Programmation des Logiciels, 12h, Level M2 IAGL, Master of Computer Science

Suivi de projets, 20h, Level M2, Master of Computer Science

Suivi d'alternants, 20h, Level M2, Master of Computer Science

Walter Rudametkin Ivey teaches at the University of Lille 1, Polytech engineering school.

GIS4 Programmation par Objets

GIS4 Architectures Logicielles

GIS2A3 (apprentissage) Projet programmation par Objet

IMA2A4 (apprentissage) Conception Modélisation Objet

GBIAAL4 Bases de données

Lionel Seinturier teaches at the University of Lille 1, IEEA faculty.

Conception d'Applications Réparties, 50h, Level M1, Master MIAGE

Conception d'Applications Réparties, 36h, Level M1, Master of Computer Science

Infrastructures et Frameworks Internet, 52h, Level M2 E-Services IAGL and TIIR, Master of Computer Science

10.2.2. Supervision

Defended HdR: Martin Monperrus, On Oracles for Automated Diagnosis and Repair of Software Bugs, 3 May 2016.

Defended PhD: Vincenzo Musco, Usages of Graphs and Synthetic Data for Software Propagation Analysis, 3 November 2016, Philippe Preux (Inria SequeL) & Martin Monperrus.

Defended PhD: Maxime Colmant, Multi-Dimensional Analysis of Software Power Consumptions in Multi-Core Architectures, 24 November 2016, Lionel Seinturier & Romain Rouvoy.

Defended PhD: Geoffrey Hecht, Détection et analyse de l'impact des défauts de code dans les applications mobiles, 30 November 2016, Romain Rouvoy.

Defended PhD: María Gómez Lacruz, Towards Improving the Quality of Mobile Apps by Leveraging Crowdsourced Feedback, 2 December 2016, Lionel Seinturier & Romain Rouvoy.

Defended PhD: Bo Zhang, Self-optimization of Infrastructure and Platform Resources in Cloud Computing, 12 December 2016, Lionel Seinturier & Romain Rouvoy.

In progress PhD: Benjamin Danglot, December 2016, Martin Monperrus.

In progress PhD: Lakhdar Meftah, November 2016, Romain Rouvoy & Isabelle Chrisment (Inria Madynes).

In progress PhD: Miguel Gonzalez, October 2016, Martin Monperrus & Romain Rouvoy & Walter Rudametkin.

In progress PhD: Sarra Habchi, Une supervision de contexte sensible à la confidentialité pour les développements logiciels en crowdsource, October 2016, Romain Rouvoy.

In progress PhD: Antoine Vastel, Cartographie de la qualité d'expérience pour l'accès à l'internet mobile, October 2016, Romain Rouvoy & Walter Rudametkin.

In progress PhD: Yahya Al Dhuraibi, Un cadre flexible pour l'élasticité dans les nuages, October 2015, Philippe Merle.

In progress PhD: Stéphanie Challita, Un cadre formel et outillé pour la gestion de toute ressource en nuage, October 2015, Philippe Merle.

In progress PhD: Thomas Durieux, Search-based Monitoring and Root Cause Diagnosis in Production, September 2015, Lionel Seinturier & Martin Monperrus.

In progress PhD: Gustavo Sousa, Towards dynamic software product lines to optimize management and reconfiguration of cloud applications, October 2012, Laurence Duchien & Walter Rudametkin Ivey.

In progress PhD: Amal Tahri, Evolution logicielle multi-vues, des réseaux domestiques au Cloud, March 2013, Laurence Duchien.

10.2.3. Juries

Laurence Duchien

HDR Tewfik Ziadi (University Paris 6), reviewer

HDR Christelle Urtado (University of Montpellier), reviewer

HDR Anne Etien (University of Lille 1), chair

HDR Martin Monperrus (University of Lille 1), garant

HDR Fabien Dagnat (University Rennes 1), reviewer

Bui Thi Mai Anh (University Paris 6), examiner

Alix Gogey (University of Lille 1), chair

Guillaume Bécan (University Rennes 1), reviewer

Martin Monperrus

Alan Charpentier (University of Bordeaux), examiner

Vincenzo Musco (University of Lille 1), co-supervisor

Romain Rouvoy

Assaad Moawad (University of Luxembourg), reviewer

Simon Dupont (École des Mines de Nantes), reviewer

Lucas Perronne (University of Grenoble), reviewer

Alexandre Caron (University of Lille 1), chair

Maxime Colmant (University of Lille 1), co-supervisor

Geoffrey Hecht (University of Lille 1 & UQAM), co-supervisor

María Gómez Lacruz (University of Lille 1), co-supervisor

Bo Zhang (University of Lille 1), co-supervisor

Lionel Seinturier

HDR Brahim Hamid (University of Toulouse), reviewer

Sebastian Martinez (University Bretagne Occidentale), reviewer

Julie Rochas (University of Nice), reviewer

Hamza Chehili (University Constantine 2, Algérie), reviewer

Naweulo Zhou (University of Grenoble), reviewer

Anca Iordache (University Rennes 1), examiner

Maxime Colmant (University of Lille 1), co-supervisor

María Gómez Lacruz (University of Lille 1), co-supervisor

Bo Zhang (University of Lille 1), co-supervisor

10.3. Popularization

Romain Rouvoy is in charge of a demonstrator of APISENSE® as part of the Xperium initiative of the LILLIAD Learning center Innovation at the University of Lille 1.

Lionel Seinturier has participated in October to the event Chercheur itinérant organized by the Inria Lille Nord Europe research center in the context of La Fête de la science. Two classrooms (3ème) have been visited. The theme of the visit was crowd-sensing and data gathering from mobile devices.

11. Bibliography

Major publications by the team in recent years

[1] B. BAUDRY, M. MONPERRUS.*The Multiple Facets of Software Diversity: Recent Developments in Year 2000 and Beyond*, in "ACM Computing Surveys", July 2015, p. 1-26, https://hal.inria.fr/hal-01182103.

[2] M. COLMANT, M. KURPICZ, P. FELBER, L. HUERTAS, R. ROUVOY, A. SOBE.*Process-level Power Estimation in VM-based Systems*, in "European Conference on Computer Systems (EuroSys)", Bordeaux, France, T. HARRIS, M. HERLIHY (editors), EuroSys'15: Proceedings of the Tenth European Conference on Computer Systems, ACM, April 2015, 14 [*DOI :* 10.1145/2741948.2741971], https://hal.inria.fr/hal-01130030.

[3] B. CORNU, E. T. BARR, L. SEINTURIER, M. MONPERRUS.*Casper: Automatic Tracking of Null Dereferences to Inception with Causality Traces*, in "Journal of Systems and Software", 2016, vol. 122, p. 52-62 [*DOI :* 10.1016/J.JSS.2016.08.062], https://hal.archives-ouvertes.fr/hal-01354090.

[4] G. HECHT, B. OMAR, R. ROUVOY, N. MOHA, L. DUCHIEN.*Tracking the Software Quality of Android Applications along their Evolution*, in "30th IEEE/ACM International Conference on Automated Software Engineering", Lincoln, Nebraska, United States, L. GRUNSKE, M. WHALEN (editors), Proceedings of the 30th IEEE/ACM International Conference on Automated Software Engineering (ASE 2015), IEEE, November 2015, 12, https://hal.inria.fr/hal-01178734.

[5] P. LAPERDRIX, W. RUDAMETKIN, B. BAUDRY.*Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints*, in "37th IEEE Symposium on Security and Privacy (S&P 2016)", San Jose, United States, May 2016, https://hal.inria.fr/hal-01285470.

[6] M. MARTINEZ, M. MONPERRUS.*Mining Software Repair Models for Reasoning on the Search Space of Automated Program Fixing*, in "Empirical Software Engineering", 2015, vol. 20, n° 1, p. 176–205 [*DOI :* 10.1007/s10664-013-9282-8], https://hal.inria.fr/hal-00903808.

[7] V. MUSCO, M. MONPERRUS, P. PREUX.*A Large-scale Study of Call Graph-based Impact Prediction using Mutation Testing*, in "Software Quality Journal", 2016 [*DOI :* 10.1007/s11219-016-9332-8], https://hal.inria.fr/hal-01346046.

[8] F. PARAISO, P. MERLE, L. SEINTURIER.*soCloud: A service-oriented component-based PaaS for managing portability, provisioning, elasticity, and high availability across multiple clouds*, in "Springer Computing", May 2016, vol. 98, n° 5, p. 539-565 [*DOI :* 10.1007/s00607-014-0421-x], https://hal.inria.fr/hal-01019385.

[9] C. QUINTON, D. ROMERO, L. DUCHIEN.*SALOON: a platform for selecting and configuring cloud environments*, in "Software: Practice and Experience", January 2016, vol. 46, p. 55-78 [*DOI :* 10.1002/SPE.2311], https://hal.inria.fr/hal-01103560.

[10] J. XUAN, M. MARTINEZ, F. DEMARCO, M. CLÉMENT, S. LAMELAS, T. DURIEUX, D. LE BERRE, M. MONPERRUS.*Nopol: Automatic Repair of Conditional Statement Bugs in Java Programs*, in "IEEE Transactions on Software Engineering", 2016 [*DOI :* 10.1109/TSE.2016.2560811], https://hal.archives-ouvertes.fr/hal-01285008.

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] M. COLMANT.*Multi-Dimensional Analysis of Software Power Consumptions in Multi-Core Architectures*, Université Lille 1 - Sciences et Technologies, November 2016, https://tel.archives-ouvertes.fr/tel-01403559.

[12] M. GOMEZ.*Towards Improving the Quality of Mobile Apps by Leveraging Crowdsourced Feedback*, Universite Lille 1 ; Inria Lille - Nord Europe, December 2016, https://tel.archives-ouvertes.fr/tel-01418298.

[13] G. HECHT.*Detection and analysis of impact of code smells in mobile applications*, Université Lille 1 : Sciences et Technologies ; Université du Québec à Montréal, November 2016, https://tel.archives-ouvertes.fr/tel-01418158.

[14] M. MONPERRUS.*On Oracles for Automated Diagnosis and Repair of Software Bugs*, Université de Lille , May 2016, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01321718.

[15] V. MUSCO.*Propagation Analysis based on Software Graphs and Synthetic Data*, Université Lille 3, November 2016, https://tel.archives-ouvertes.fr/tel-01398903.

[16] B. ZHANG.*Self-optimization of Infrastructure and Platform Resources in Cloud Computing*, Lille1, December 2016, https://tel.archives-ouvertes.fr/tel-01417289.

Articles in International Peer-Reviewed Journal

[17] B. CORNU, E. T. BARR, L. SEINTURIER, M. MONPERRUS.*Casper: Automatic Tracking of Null Dereferences to Inception with Causality Traces*, in "Journal of Systems and Software", 2016, vol. 122, p. 52-62 [*DOI :* 10.1016/J.JSS.2016.08.062], https://hal.archives-ouvertes.fr/hal-01354090.

[18] I. GONZALEZ-HERRERA, J. BOURCIER, E. DAUBERT, W. RUDAMETKIN, O. BARAIS, F. FOUQUET, J.-M. JÉZÉQUEL, B. BAUDRY.*ScapeGoat: Spotting abnormal resource usage in component-based reconfigurable software systems*, in "Journal of Systems and Software", 2016 [*DOI :* 10.1016/J.JSS.2016.02.027], https://hal.inria.fr/hal-01354999.

[19] M. GÓMEZ, B. ADAMS, M. WALID, M. MONPERRUS, R. ROUVOY.*App Store 2.0: From Crowd Information to Actionable Feedback in Mobile Ecosystems*, in "IEEE Software", March 2017, 6, https://hal.inria.fr/hal-01406422.

[20] M. MARTINEZ, T. DURIEUX, R. SOMMERARD, J. XUAN, M. MONPERRUS.*Automatic Repair of Real Bugs in Java: A Large-Scale Experiment on the Defects4J Dataset*, in "Empirical Software Engineering", 2016 [*DOI :* 10.1007/S10664-016-9470-4], https://hal.archives-ouvertes.fr/hal-01387556.

[21] V. MUSCO, M. MONPERRUS, P. PREUX.*A Large-scale Study of Call Graph-based Impact Prediction using Mutation Testing*, in "Software Quality Journal", 2016 [*DOI :* 10.1007/S11219-016-9332-8], https://hal.inria.fr/hal-01346046.

[22] F. PARAISO, P. MERLE, L. SEINTURIER.*soCloud: A service-oriented component-based PaaS for managing portability, provisioning, elasticity, and high availability across multiple clouds*, in "Springer Computing", May 2016, vol. 98, n^{o} 5, p. 539-565 [*DOI :* 10.1007/S00607-014-0421-X], https://hal.inria.fr/hal-01019385.

[23] C. QUINTON, D. ROMERO, L. DUCHIEN.*SALOON: a platform for selecting and configuring cloud environments*, in "Software: Practice and Experience", January 2016, vol. 46, p. 55-78 [*DOI :* 10.1002/SPE.2311], https://hal.inria.fr/hal-01103560.

[24] J. XUAN, B. CORNU, M. MARTINEZ, B. BAUDRY, L. SEINTURIER, M. MONPERRUS.*B-Refactoring: Automatic Test Code Refactoring to Improve Dynamic Analysis*, in "Information and Software Technology", 2016, vol. 76, p. 65-80 [*DOI :* 10.1016/J.INFSOF.2016.04.016], https://hal.archives-ouvertes.fr/hal-01309004.

[25] J. XUAN, M. MARTINEZ, F. DEMARCO, M. CLÉMENT, S. LAMELAS, T. DURIEUX, D. LE BERRE, M. MONPERRUS.*Nopol: Automatic Repair of Conditional Statement Bugs in Java Programs*, in "IEEE Transactions on Software Engineering", 2016 [*DOI :* 10.1109/TSE.2016.2560811], https://hal.archives-ouvertes.fr/hal-01285008.

International Conferences with Proceedings

[26] A. CARETTE, M. A. AIT YOUNES, G. HECHT, N. MOHA, R. ROUVOY.*Investigating the Energy Impact of Android Smells*, in "24th International IEEE Conference on Software Analysis, Evolution and Reengineering (SANER)", Klagenfurt, Austria, A. MARCUS, G. BAVOTA (editors), Proceedings of the 24th International IEEE Conference on Software Analysis, Evolution and Reengineering (SANER), IEEE, February 2017, 10, https://hal.inria.fr/hal-01403485.

[27] P. FABIO, P. MERLE, N. MOHA, Y.-G. GUÉHÉNEUC.*Are REST APIs for Cloud Computing Well-Designed? An Exploratory Study*, in "14th International Conference on Service Oriented Computing (ICSOC)", Banff, Canada, October 2016, https://hal.archives-ouvertes.fr/hal-01352738.

[28] M. GOMEZ, R. ROUVOY, B. ADAMS, L. SEINTURIER.*Mining Test Repositories for Automatic Detection of UI Performance Regressions in Android Apps*, in "13th International Conference on Mining Software Repositories (MSR'16)", Austin, Texas, United States, R. ROBBES, C. BIRD (editors), Proceedings of the 13th International Conference on Mining Software Repositories, IEEE, May 2016, https://hal.inria.fr/hal-01280784.

[29] I. GONZALEZ-HERRERA, J. BOURCIER, W. RUDAMETKIN, O. BARAIS, F. FOUQUET.*Squirrel: Architecture Driven Resource Management*, in "SAC - 31st Annual ACM Symposium on Applied Computing", Pisa, Italy, April 2016 [*DOI :* 10.1145/0000000.0000000], https://hal.inria.fr/hal-01355000.

[30] M. GÓMEZ, R. ROUVOY, B. ADAMS, L. SEINTURIER.*Reproducing Context-sensitive Crashes of Mobile Apps using Crowdsourced Monitoring*, in "IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft'16)", Austin, Texas, United States, L. FLYNN, P. INVERARDI (editors), Proceedings of the 3rd IEEE/ACM International Conference on Mobile Software Engineering and Systems, IEEE, May 2016, https://hal.inria.fr/hal-01276926.

[31] A. HAVET, V. SCHIAVONI, P. FELBER, M. COLMANT, R. ROUVOY, C. FETZER.*GenPack: A Generational Scheduler for Cloud Data Centers*, in "5th IEEE International Conference on Cloud Engineering (IC2E)", Vancouver, Canada, I. GUPTA, J. LIU (editors), Proceedings of the 5th IEEE International Conference on Cloud Engineering (IC2E), IEEE, April 2017, 10, https://hal.inria.fr/hal-01403486.

[32] G. HECHT, N. MOHA, R. ROUVOY.*An Empirical Study of the Performance Impacts of Android Code Smells*, in "IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft'16)", Austin, Texas, United States, L. FLYNN, P. INVERARDI (editors), Proceedings of the 3rd IEEE/ACM International Conference on Mobile Software Engineering and Systems, IEEE, May 2016, vol. 1, n^o 1, https://hal.inria.fr/hal-01276904.

[33] P. LAPERDRIX, W. RUDAMETKIN, B. BAUDRY.*Beauty and the Beast: Diverting modern web browsers to build unique browser fingerprints*, in "37th IEEE Symposium on Security and Privacy (S&P 2016)", San Jose, United States, May 2016, https://hal.inria.fr/hal-01285470.

[34] M. MARTINEZ, M. MONPERRUS.*ASTOR: A Program Repair Library for Java*, in "Proceedings of ISSTA, Demonstration Track", Saarbrücken, Germany, 2016, p. 441-444 [*DOI :* 10.1145/2931037.2948705], https://hal.archives-ouvertes.fr/hal-01321615.

[35] V. MUSCO, A. CARETTE, M. MONPERRUS, P. PREUX.*A Learning Algorithm for Change Impact Prediction*, in "5th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering", Austin, United States, May 2016, https://hal.inria.fr/hal-01279620.

[36] V. MUSCO, M. MONPERRUS, P. PREUX.*Mutation-Based Graph Inference for Fault Localization*, in "International Working Conference on Source Code Analysis and Manipulation", Raleigh, United States, October 2016, https://hal.inria.fr/hal-01350515.

[37] F. PARAISO, S. CHALLITA, Y. AL-DHURAIBI, P. MERLE.*Model-Driven Management of Docker Containers*, in "9th IEEE International Conference on Cloud Computing (CLOUD)", San Francisco, United States, June 2016, 8, https://hal.inria.fr/hal-01314827.

[38] G. SOUSA, W. RUDAMETKIN, L. DUCHIEN.*Automated Setup of Multi-Cloud Environments for Microservices-Based Applications*, in "9th IEEE International Conference on Cloud Computing", San Francisco, United States, June 2016, https://hal.inria.fr/hal-01312606.

[39] G. SOUSA, W. RUDAMETKIN, L. DUCHIEN.*Extending Feature Models with Relative Cardinalities*, in "20th International Systems and Software Product Line Conference", Beijing, China, September 2016, https://hal.inria.fr/hal-01312751.

[40] B. ZHANG, Y. AL-DHURAIBI, F. PARAISO, R. ROUVOY, L. SEINTURIER.*CloudGC: Recycling Idle Virtual Machines in the Cloud*, in "5th IEEE International Conference on Cloud Engineering (IC2E)", Vancouver, Canada, I. GUPTA, J. LIU (editors), Proceedings of the 5th IEEE International Conference on Cloud Engineering (IC2E), IEEE, April 2017, 10, https://hal.inria.fr/hal-01403488.

[41] B. ZHANG, F. KŘIKAVA, R. ROUVOY, L. SEINTURIER.*Self-Balancing Job Parallelism and Throughput in Hadoop*, in "16th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS)", Heraklion, Crete, Greece, E. KALYVIANAKI, M. JELASITY (editors), Proceedings of the 16th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS), Springer, June 2016, vol. 9687, p. 129-143 [*DOI :* 10.1007/978-3-319-39577-7_11], https://hal.inria.fr/hal-01294834.

National Conferences with Proceeding

[42] R. SOMMERARD, R. ROUVOY.*Towards Privacy-Preserving Data Dissemination in Crowd-Sensing Middleware Platform*, in "11èmes journées francophones Mobilité et Ubiquité (UbiMob'16)", Lorient, France, July 2016, 6, https://hal.inria.fr/hal-01332588.

Conferences without Proceedings

[43] T. DURIEUX, B. CORNU, L. SEINTURIER, M. MONPERRUS.*Dynamic Patch Generation for Null Pointer Exceptions Using Metaprogramming*, in "IEEE International Conference on Software Analysis, Evolution and Reengineering", Klagenfurt, Austria, February 2017, https://hal.archives-ouvertes.fr/hal-01419861.

[44] T. DURIEUX, M. MONPERRUS.*DynaMoth: Dynamic Code Synthesis for Automatic Program Repair*, in "11th International Workshop in Automation of Software Test (AST 2016)", Austin, United States, May 2016 [*DOI :* 10.1145/2896921.2896931], https://hal.archives-ouvertes.fr/hal-01279233.

[45] G. SOUSA, W. RUDAMETKIN, L. DUCHIEN.*Software Product Lines for Multi-Cloud Microservices-Based Applications*, in "6th International Workshop on Cloud Data and Platforms", Londres, United Kingdom, April 2016, https://hal.inria.fr/hal-01302184.

[46] G. SOUSA.*Software Product Lines for Multi-Cloud Microservices Configuration*, in "Journées Cloud GdR RSD", Nice, France, September 2016, https://hal.inria.fr/hal-01355864.

Scientific Books (or Scientific Book chapters)

[47] F. KRIKAVA, P. COLLET, R. ROUVOY, L. SEINTURIER.*Contracts-based Control Integration into Software Systems*, in "Software Engineering for Self-Adaptive Systems 3", R. DE LEMOS, D. GARLAN, C. GHEZZI, H. GIESE (editors), Springer, February 2016, https://hal.inria.fr/hal-01257205.

[48] M. LITOIU, M. SHAW, G. TAMURA, N. M. VILLEGAS, H. MÜLLER, H. GIESE, E. RUTTEN, R. ROUVOY.*What Can Control Theory Teach Us About Assurances in Self-Adaptive Software Systems?*, in "Software Engineering for Self-Adaptive Systems 3: Assurances", R. DE LEMOS, D. GARLAN, C. GHEZZI, H. GIESE (editors), Springer, February 2016, https://hal.inria.fr/hal-01281063.

Research Reports

[49] B. DANGLOT, P. PREUX, B. BAUDRY, M. MONPERRUS.*Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation*, HAL, 2016, n^o hal-01378523, https://hal.archives-ouvertes.fr/hal-01378523.

[50] T. DURIEUX, M. MONPERRUS.*IntroClassJava: A Benchmark of 297 Small and Buggy Java Programs*, Universite Lille 1, 2016, https://hal.archives-ouvertes.fr/hal-01272126.

[51] G. SOUSA, W. RUDAMETKIN, L. DUCHIEN.*Extending Feature Models with Relative Cardinalities*, Université Lille 1 ; CRIStAL UMR 9189 ; Inria Lille - Nord Europe, January 2016, n^o 8843, 24, https://hal.inria.fr/hal-01257909.

Other Publications

[52] C. DORFFER, M. PUIGT, G. DELMAIRE, G. ROUSSEL, R. ROUVOY.*Traitement de signaux irrégulièrement échantillonnés issus du mobile crowd-sensing pour la surveillance de la qualité de l'air*, October 2016, Innovatives Big Data, Poster, https://hal.archives-ouvertes.fr/hal-01371240.

[53] T. DURIEUX, Y. HAMADI, M. MONPERRUS.*BanditRepair: Speculative Exploration of Runtime Patches*, March 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01293596.

[54] T. DURIEUX, Y. HAMADI, M. MONPERRUS.*Production-Driven Patch Generation and Validation*, September 2016, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01370709.

[55] R. ROUVOY.*APISENSE®: Mobile crowd-sensing made easy!*, June 2016, Keynote of ECAAS'16, https://hal.inria.fr/hal-01332594.

References in notes

[56] L. CAPRA, W. EMMERICH, C. MASCOLO.*A micro-economic approach to conflict resolution in mobile computing*, in "SIGSOFT FSE", 2002, p. 31-40.

[57] B. CORNU.*Automatic Analysis and Repair of Exception Bugs for Java Programs*, Université de Lille, November 2015, https://tel.archives-ouvertes.fr/tel-01250092.

[58] V. DALLMEIER, A. ZELLER, B. MEYER.*Generating Fixes from Object Behavior Anomalies*, in "Proceedings of the 2009 IEEE/ACM International Conference on Automated Software Engineering", Washington, DC, USA, ASE '09, IEEE Computer Society, 2009, p. 550–554, http://dx.doi.org/10.1109/ASE.2009.15.

[59] F. FOUQUET, E. DAUBERT, N. PLOUZEAU, O. BARAIS, J. BOURCIER, J.-M. JÉZÉQUEL.*Dissemination of Reconfiguration Policies on Mesh Networks*, in "Proceedings of 12th IFIP International Conference on Distributed Applications and Interoperable Systems (DAIS'12)", Stockholm, Sweden, K. M. GÖSCHKA, S. HARIDI (editors), Lecture Notes in Computer Science, Springer, June 2012, vol. 7272, p. 16-30.

[60] M. GOLDMANN, G. KREITZ.*Measurements on the spotify peer-assisted music-on-demand streaming system*, in "Peer-to-Peer Computing", T. ASAMI, T. HIGASHINO (editors), IEEE, 2011, p. 206-211.

[61] M. HALL, E. FRANK, G. HOLMES, B. PFAHRINGER, P. REUTEMANN, I. H. WITTEN.*The WEKA data mining software: an update*, in "SIGKDD Explorer Newsletter", 2009, vol. 11, n^{o} 1, p. 10–18.

[62] M. HARMAN.*Automated patching techniques: the fix is in*, in "Communications of the ACM", May 2010, vol. 53, n^{o} 5, p. 108–108, http://doi.acm.org/10.1145/1735223.1735248.

[63] V. ISSARNY, N. GEORGANTAS, S. HACHEM, A. ZARRAS, P. VASSILIADIS, M. AUTILI, M. A. GEROSA, A. BEN HAMIDA.*Service-Oriented Middleware for the Future Internet: State of the Art and Research Directions*, in "Journal of Internet Services and Applications", May 2011, vol. 2, n^{o} 1, p. 23-45, http://dx.doi.org/10.1007/s13174-011-0021-3.

[64] W. LI, S. HU, J. LI, H.-A. JACOBSEN.*Community Clustering for Distributed Publish/Subscribe Systems*, in "CLUSTER", IEEE, 2012, p. 81-89.

[65] M. LÉGER, T. LEDOUX, T. COUPAYE.*Reliable Dynamic Reconfigurations in a Reflective Component Model*, in "CBSE", L. GRUNSKE, R. REUSSNER, F. PLASIL (editors), Lecture Notes in Computer Science, Springer, 2010, vol. 6092, p. 74-92.

[66] M. MARTINEZ.*Extraction and Analysis of Knowledge for Automatic Software Repair*, Université Lille 1, October 2014, https://hal.archives-ouvertes.fr/tel-01078911.

[67] L. PROVENSI, F. ELIASSEN, R. VITENBERG, R. ROUVOY.*Improving Context Interpretation by Using Fuzzy Policies: The Case of Adaptive Video Streaming*, in "28th ACM Symposium on Applied Computing (SAC) - 8th Track on Dependable and Adaptive Distributed Systems (DADS)", Coimbra, Portugal, K. M. GÖSCHKA, R. OLIVEIRA, P. PIETZUCH, G. RUSSELLO (editors), ACM, March 2013, vol. 1, p. 415-422, Best paper award, http://hal.inria.fr/hal-00799136.

[68] M. RINARD, C. CADAR, D. DUMITRAN, D. ROY, T. LEU, W. BEEBEE JR.*Enhancing server availability and security through failure-oblivious computing*, in "Proceedings of the 6th conference on Symposium on Operating Systems Design & Implementation (OSDI'06)", USENIX Association, 2004, p. 21–21.

[69] M. SALEHIE, L. TAHVILDARI.*Self-adaptive software: Landscape and research challenges*, in "ACM Transactions on Autonomous and Adaptive Systems", May 2009, vol. 4, n^{o} 2, p. 14:1–14:42, http://doi.acm.org/10.1145/1516533.1516538.

[70] H. SCHAFFERS, M. SHARPE.*Services in the Future Internet*, April 2011, FP8 Expert Group, EU Directorate-General Information Society & Media, http://cordis.europa.eu/fp7/ict/ssai/docs/softwareconsultationreportfeb2011.pdf.

[71] L. SCHUBERT, K. JEFFERY, B. NEIDECKER-LUTZ.*A Roadmap for Advanced Cloud Technologies under H2020*, Dec 2012, http://cordis.europa.eu/fp7/ict/ssai/docs/cloud-expert-group/roadmap-dec2012-vfinal.pdf.

[72] L. SEINTURIER, P. MERLE, R. ROUVOY, D. ROMERO, V. SCHIAVONI, J.-B. STEFANI.*A Component-Based Middleware Platform for Reconfigurable Service-Oriented Architectures*, in "Software: Practice and Experience", May 2012, vol. 42, n^{o} 5, p. 559-583 [*DOI :* 10.1002/SPE.1077], http://hal.inria.fr/inria-00567442.

[73] J. URBAN.*NESSI Research Priorities for the next Framework Programme for Research and Technological Development FP8*, May 2011, http://www.nessi-europe.eu/files/Docs/NESSI%20SRA_update_May_2011_V1-0.pdf.

[74] Y. WEI, Y. PEI, C. A. FURIA, L. S. SILVA, S. BUCHHOLZ, B. MEYER, A. ZELLER.*Automated fixing of programs with contracts*, in "Proceedings of the 19th international symposium on Software testing and analysis", New York, NY, USA, ISSTA '10, ACM, 2010, p. 61–72, http://doi.acm.org/10.1145/1831708.1831716.

[75] C. WEISS, R. PREMRAJ, T. ZIMMERMANN, A. ZELLER.*How long will it take to fix this bug?*, in "Proceedings of the Fourth International Workshop on Mining Software Repositories", IEEE Computer Society, 2007, 1.

[76] T. XIE, S. THUMMALAPENTA, D. LO, C. LIU.*Data Mining for Software Engineering*, in "IEEE Computer", August 2009, vol. 42, n^{o} 8, p. 55-62, http://doi.ieeecomputersociety.org/10.1109/MC.2009.256.

[77] W. XU, L. HUANG, A. FOX, D. A. PATTERSON, M. I. JORDAN.*Detecting large-scale system problems by mining console logs*, in "Proceedings of the 22nd ACM Symposium on Operating Systems Principles (SOSP'09)", Big Sky, Montana, USA, J. N. MATTHEWS, T. E. ANDERSON (editors), ACM, October 2009, p. 117-132.

[78] R. DE LEMOS, H. GIESE, H. A. MULLER, M. SHAW, J. ANDERSSON, L. BARESI, B. BECKER, N. BENCOMO, Y. BRUN, B. CUKIC, R. DESMARAIS, S. DUSTDAR, G. ENGELS, K. GEIHS, K. M. GÖSCHKA, A. GORLA, V. GRASSI, P. INVERARDI, G. KARSAI, J. KRAMER, M. LITOIU, A. LOPES, J. MAGEE, S. MALEK, S. MANKOVSKII, R. MIRANDOLA, J. MYLOPOULOS, O. NIERSTRASZ, M. PEZZÈ, C. PREHOFE, W. SCHÄFER, R. SCHLICHTING, B. SCHMERL, D. B. SMITH, J. P. SOUSA, G. TAMURA, L. TAHVILDARI, N. M. VILLEGAS, T. VOGEL, D. WEYNS, K. WONG, J. WUTTKE.*Software Engineering for Self-Adaptive Systems: A Second Research Roadmap*, in "Software Engineering for Self-Adaptive Systems", R. DE LEMOS, H. GIESE, H. MÜLLER, M. SHAW (editors), Dagstuhl Seminar Proceedings, Springer, 2013, vol. 7475, p. 1-26, http://hal.inria.fr/inria-00638157.