Activity Report Lille - Nord Europe 2018

Edition: 2019-03-11

List of Inria's Research Teams

1. Project-Team BONSAI 4
2. Team BONUS 21
3. Project-Team DEFROST 50
4. Project-Team FUN 73
5. Team GAIA 108
6. Project-Team INOCS 134
7. Project-Team LINKS 161
8. Team LOKI 181
9. Project-Team MAGNET 210
10. Team MEPHYSTO-POST 237
11. Project-Team MODAL 252
12. Team NON-A POST 288
13. Project-Team RAPSODI 305
14. Project-Team RMOD 334
15. Project-Team SEQUEL 360
16. Project-Team SPIRALS 401

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

1. Team, Visitors, External Collaborators 7
2. Overall Objectives 7
3. Research Program 8
- 3.1. Sequence processing for Next Generation Sequencing 8
- 3.2. Noncoding RNA 9
- 3.3. Genome structures 9
- 3.4. Nonribosomal peptides 9

4. Highlights of the Year 9
5. New Software and Platforms 10
- 5.1. BCALM 2 10
- 5.2. NORINE 10
- 5.3. Vidjil 11
- 5.4. MATAM 11

6. New Results 12
- 6.1. Exploration of transcriptomes 12
- 6.2. Modeling of alternative transcripts with long reads 12
- 6.3. Read against read comparison for Nanopore data 12
- 6.4. Annotation of the OC43 coronavirus genome 13
- 6.5. Small RNAs catalog in oilseed rape 13
- 6.6. Identifying systematic sequencing errors 13
- 6.7. Indexing labelled sequences 13
- 6.8. Tree representations 14
- 6.9. Co-linear chaining on graphs 14
- 6.10. Representations of de Bruijn graphs 14
- 6.11. Readability of overlap graphs 14
- 6.12. Nonribosomal peptides 14

7. Partnerships and Cooperations 15
- 7.1. National Initiatives 15
  - 7.1.1. ANR 15
  - 7.1.2. ADT 15
- 7.2. European Initiatives 15
- 7.3. International Research Visitors 15

8. Dissemination 16
- 8.1. Promoting Scientific Activities 16
  - 8.1.1. Scientific Events Organisation 16
  - 8.1.2. Scientific Events Selection 16
    - 8.1.2.1. Chair of Conference Program Committees 16
    - 8.1.2.2. Member of the Conference Program Committees 16
    - 8.1.2.3. Reviewer 16
  - 8.1.3. Journal 16
    - 8.1.3.1. Member of the Editorial Boards 16
    - 8.1.3.2. Reviewer - Reviewing Activities 16
  - 8.1.4. Scientific Expertise 16
  - 8.1.5. Research Administration 16
- 8.2. Teaching - Supervision - Juries 17
  - 8.2.1. Teaching 17
  - 8.2.2. Teaching administration 17
  - 8.2.3. Supervision 17
  - 8.2.4. Juries 18

8.3. Popularization 18
8.3.1. Internal or external Inria responsibilities 18
8.3.2. Education 18
9. Bibliography **18**

Project-Team BONSAI

Creation of the Project-Team: 2011 January 01, end of the Project-Team: 2018 December 31

Keywords:

Computer Science and Digital Science:

A7.1. - Algorithms
A8.1. - Discrete mathematics, combinatorics
A8.7. - Graph theory

Other Research Topics and Application Domains:

B1.1.5. - Immunology
B1.1.6. - Evolutionnary biology
B1.1.7. - Bioinformatics
B1.1.11. - Plant Biology
B2.2.3. - Cancer

1. Team, Visitors, External Collaborators

Research Scientists

Hélène Touzet [Team leader, CNRS, Senior Researcher, HDR]
Samuel Blanquart [Inria, Researcher, until Apr 2018]
Rayan Chikhi [CNRS, Researcher]

Faculty Members

Stéphane Janot [Université de Lille, Associate Professor]
Laurent Noé [Université de Lille, Associate Professor]
Maude Pupin [Université de Lille, Associate Professor, HDR]
Mikaël Salson [Université de Lille, Associate Professor]
Jean-Stéphane Varré [Université de Lille, Professor, HDR]

Technical Staff

Areski Flissi [CNRS]
Ryan Herbert [Inria, until Mar 2018]
Mael Kerbiriou [Inria]

PhD Students

Quentin Bonenfant [CNRS]
Pierre Marijon [Inria]
Chadi Saad [Université de Lille, until Aug 2018]

Post-Doctoral Fellow

Aymeric Antoine-Lorquin [Inria]

Administrative Assistant

Amelie 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. Highlights of the Year

4.1. Highlights of the Year

The team was actively involved in the organization of the international conference RECOMB in Paris (April 2018), that was attended by more than 800 people.

4.1.1. Awards

First place at the metagenomics assembly challenge organized by the company Mosaic DNANexus: https://www.businesswire.com/news/home/20180620005408/en/DNAnexus-Powered-Mosaic-Microbiome-Platform-Announces-Winners-Community

5. New Software and Platforms

5.1. 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: Antoine Limasset, Paul Medvedev and Rayan Chikhi
- Contact: Rayan Chikhi
- Publication: Compacting de Bruijn graphs from sequencing data quickly and in low memory
- URL: https://github.com/GATB/bcalm

5.2. NORINE

Nonribosomal peptides resource
KEYWORDS: Drug development - Knowledge database - Chemistry - Graph algorithmics - Genomics - Biology - Biotechnology - Bioinformatics - 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.
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 easily submit new nonribosomal peptides, Smiles2monomers (s2m), a tool that deciphers the monomeric structure of polymers from their chemical structure.

- Participants: Areski Flissi, Juraj Michalik, Laurent Noé, Maude Pupin, Stéphane Janot, Valerie Leclère and Yoann Dufresne
- Partners: CNRS - Université Lille 1 - Institut Charles Viollette
- Contact: Maude Pupin
- Publications: Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing - Smiles2Monomers: a link between chemical and biological structures for polymers - Norine: a powerful resource for novel nonribosomal peptide discovery - NORINE: a database of nonribosomal peptides. - Bioinformatics Tools for the Discovery of New Nonribosomal Peptides
- URL: http://bioinfo.lille.inria.fr/NRP

5.3. Vidjil

High-Throughput Analysis of V(D)J Immune Repertoire
KEYWORDS: Cancer - Indexation - NGS - Bioinformatics - Drug development
SCIENTIFIC DESCRIPTION: Vidjil is made of three components: an algorithm, a visualization 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 gathers 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: Florian Thonier, Marc Duez, Mathieu Giraud, Mikaël Salson, Ryan Herbert and Tatiana Rocher
- Partners: CNRS - Inria - Université de Lille - CHRU Lille
- Contact: Mathieu Giraud
- Publications: High-Throughput Immunogenetics for Clinical and Research Applications in Immunohematology: Potential and Challenges. - High-throughput sequencing in acute lymphoblastic leukemia: Follow-up of minimal residual disease and emergence of new clones - Diagnostic et suivi des leucémies aiguës lymphoblastiques (LAL) par séquençage haut-débit (HTS) - Multiclonal Diagnosis and MRD Follow-up in ALL with HTS Coupled with a Bioinformatic Analysis - A dataset of sequences with manually curated V(D)J designations - Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing - Multi-loci diagnosis of acute lymphoblastic leukaemia with high-throughput sequencing and bioinformatics analysis - Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing - The predictive strength of next-generation sequencing MRD detection for relapse compared with current methods in childhood ALL.
- URL: http://www.vidjil.org

5.4. MATAM

Mapping-Assisted Targeted-Assembly for Metagenomics
KEYWORDS: Metagenomics - Genome assembling - Graph algorithmics
SCIENTIFIC DESCRIPTION: MATAM relies on the construction of a read overlap graph. Overlaps are computed using SortMeRNA. The overlap graph is simplified into relevant components related to specific and conserved regions. Components are assembled into contigs using SGA and contigs are finally assembled into scaffolds. The process yields nearly full length marker sequences with a very low error rate compared to the state of the art approaches. Taxonomic assignation of the obtained scaffolds is performed using the RDP classifier and is represented using Krona.

FUNCTIONAL DESCRIPTION: MATAM provides targeted genes assembly from the short metagenomic reads issued from environmental samples sequencing. Its default application focuses on the gold standard for species identification, 16S / 18S ribosomal RNA SSU genes. The produced gene scaffolds are highly accurate and suitable for precise taxonomic assignation. The software also provides a RDP classification for the reconstructed scaffolds as well as an estimation of the relative population sizes.

- Participants: Hélène Touzet, Pierre Pericard, Yoann Dufresne, Samuel Blanquart and Loïc Couderc
- Contact: Hélène Touzet
- Publication: MATAM: reconstruction of phylogenetic marker genes from short sequencing reads in metagenomes
- URL: https://github.com/bonsai-team/matam

6. New Results

6.1. Exploration of transcriptomes

In 2016 we produced a method called CG-Alcode able to compare transcripts repertoires of a given pair of orthologous genes. We applied our method to compare human and mouse transcriptomes. This year, in collaboration with C.Belleannée (DYLISS, Inria Rennes) we explored the comparison of multiple species. We inspected human, mouse and dog transcriptomes. We thus were able to predict a large number of putative transcripts in both human, mouse and dog based on known transcripts. Those results allow to investigate which functional sites are conserved and which genes have the same set of transcripts (known or putative).

6.2. Modeling of alternative transcripts with long reads

In the context of transcriptomic analyses based on third generation sequencing data (ONT), we started to explore the following problem : given a transcriptomic experiment, a gene of interest, select reads related to the given gene and find exon junctions. As we have done in the CG-Alcode project, we aim to model the gene as an alphabet of exonic blocks, each transcript being a word over this alphabet. This work takes place in the context of ANR ASTER for which we deal with mouse transcriptomic data in brain and liver. Built models will allow to query human genes to discover putative transcripts.

6.3. Read against read comparison for Nanopore data

In the team, we developed two years ago seeds with errors, which allow to find all common approximate patterns with a limited number of errors. The idea behind these seeds, called 01^*0 seeds, is to divide the sequence in blocks so that the distribution of errors is no longer random. This year, we have used these seeds in the context of long reads analysis. With this data, reads against reads comparison suffers from a high loss of sensitivity, because the single *read error-rate* is already high. Our application case is the detection of adapter sequences in ONT sequencing. We have shown that the use of these seeds instead of exact k-mers allowed a more accurate reconstruction of the sequences of the adapters. The method takes two steps: first the identification of k-mers potentially composing the adapter using a counting approach that takes into account errors in the read, and then the reconstruction of the complete sequence of the adapter with a greedy algorithm. Our results show that the seeds with errors allow to obtain accurate consensus sequences for more 80% of the samples, compared to 40% with the usual k-mer approach. This work was done within the ANR ASTER during the first year of the thesis of Quentin Bonenfant and was presented at the national workshop Seqbio 2018.

6.4. Annotation of the OC43 coronavirus genome

OC43 coronavirus is recognized as frequent cause of respiratory infection. We have conducted a bioinformatics study of 8 coronavirus genomes collected from patients at Lille hospital : gene annotation, phylogenetic analysis and amino acids substitutions. Several genotypes (B, E, F and G) were identified and two clusters of patients were defined from chronological data and phylogenetic trees based on the genomic sequences,. Analyses of amino acids substitutions of the S protein sequences identify substitutions specific of genotype F strains circulating among French people. This work is a collaboration with Anne Goffard (CHRU Lille and CIIL).

6.5. Small RNAs catalog in oilseed rape

Polyploidy – and notably allopolyploidy that involves interspecific hybridization – has played a major role in the evolution of plants, partly because this process is often associated with genomic structure and expression changes. Homeologous exchanges (HE) – i.e. between the constituent subgenomes– have been demonstrated to be frequent in allopolyploids and could be involved in the origin and maintenance of polyploids. While their influence on gene content has poorly been studied until recently, little is known about their impact on gene expression. Together with K. Alix (Inra Moulon), we have analyzed the impact of HEs that have been characterized in resynthesized oilseed rapes, on the repertoire of micro RNAs. Our main objective was to assess the relations that could exist between structural variation and modifications of gene expression through changes in miRNA regulation. The analysis was based on the small RNA-seq catalog obtained with the bioinformatic tool miRkwood, developed in BONSAI. We have built a microRNA database for the diploid subgenomes AA from *Brassica rapa* and CC for *Brassica oleracea* that correspond to the progenitors of the resynthesized *Brassica napus* allotetraploids (AACC). Integrating miRNA prediction and genomic location of HEs allowed us to infer relationships between microRNA restructuring and non-additivity of gene expression in polyploid hybrids.

6.6. Identifying systematic sequencing errors

Discovering over-represented approximate motifs in DNA sequences is an essential part of bioinformatics, which has been studied extensively. However, it remains a difficult challenge, especially with the huge quantity of data generated by high throughput sequencing technologies. We have developed an exact discriminative method for IUPAC motifs discovery in large sets of DNA sequences. The approach uses mutual information (MI) as an objective function to search for over-represented degenerate motifs in a lattice [7].

The algorithm was applied to the problem of *Sequence-Specific Errors*. Next Generation Sequencing, and further Single-Molecule Sequencing technologies are known to produce a highly variable error rate. A common method to overcome these sequencing errors is to increase the *coverage*. However, Sequence-Specific Errors are recurrent errors that depend on the upstream nucleotidic context, and can thus be confused with true genomic variations when the read coverage increases. Our algorithm was able to find motifs associated to sequencing errors and therefore to improve variant calling. This method has also tested on ChIP-seq datasets, and compared with five state-of-the art methods, where it was experimentally shown to perform as well as the best one, while be resistant to down-sampling.

This work was done during the thesis of Chadi Saad, and as a collaboration with Martin Figeac (Univ. Lille - Plateau de génomique fonctionnelle et structurale), Julie Leclerc and Marie-Pierre Buisine (CHRU de Lille - JPARC), and Hugues Richard (Sorbonne Université - Laboratory Computational and Quantitative Biology).

6.7. Indexing labelled sequences

We designed a compressed full-text index structure able to index a whole text with labels attached to every letter in the text [6]. This work will be applied to DNA sequences and more precisely V(D)J recombinations which are complex genomic rearrangements occurring in lymphocytes. The index will be used to index labelled V(D)J recombinations, which are labelled with their V, D and J gene. As the index we conceived is scalable, we will index V(D)J recombinations from thousands of samples and give access to this data through the Vidjil platform.

6.8. Tree representations

We found an intriguing duality between two well-known representations of trees [12]. This work concerns data structures and succinct tree representations. The Balanced Parenthesis representation of trees consists of encoding the structure of any tree using a series of opening and closing parentheses. The DFUDS representation is similar, but differs in how each node is encoded (also using parentheses). By relating both BP and DFUDS representations, we obtained improvements for a basic fundamental problem: the Minimum Length Interval Query problem. We also reported unnoted commonalities in recent solution to the Range Minimum Query problem.

6.9. Co-linear chaining on graphs

We reported the first algorithm that perform co-linear chaining between a sequence and a directed acyclic graph (DAG) [9]. This work concerns dynamic programming algorithms and sequencing alignment. The problem of co-linear chaining is a classical bioinformatics problem, which has immediate application to sequence alignment, as it is used as a filter remove spurious alignment seeds. Co-linear chaining is typically solved using a simple dynamic programming algorithm. Yet, representations of genomes using graphs instead of sequences have recently became an active research topic. As a results, the problem of aligning a sequence to a sequence graph merits consideration. This work provides the first step towards tackling practical sequence-to-graph alignment instances, by first considering the case when the graph is a DAG. We designed a $O(k|E|log|V|)$ algorithm to solve co-linear chaining on DAGs, which matches the optimal solution for the classical sequence variant, i.e. when the graph is a path.

6.10. Representations of de Bruijn graphs

We designed the first practical data structure for representing large de Bruijn graphs, which supports insertions and deletions of nodes [3]. This work concerns *de novo* assembly and several other k-mer-related bioinformatics problems. The representation of de Bruijn graphs is a transversal bioinformatics question that has enjoyed recent applications in genome, metagenome and transcriptome assembly and quantification. To this date, efficient data structures were essentially static. In this work we provided an implementation of a dynamic data structure that combines perfect hashing, Karp-Rabin hashing, and forests. Practical tests show that this structure is highly competitive with the state of the art.

6.11. Readability of overlap graphs

We report further progress on the study of a theoretical parameter of graph named *readability* [8]. This work concerns graph theory mainly. The readability parameter measures the minimal length of strings that would be needed in order to label a graph such that it is an overlap graph over a set of strings of that length. So far, recent works on readability have not elucidated many aspects related to this parameter: the complexity of computing it is open, and it is not even known whether the corresponding decision problem is in NP. The only upper bound known for this parameter is exponential. This work focuses on certain graph families: bipartite chain graphs, grids, induced subgraphs of grids, and provides a characterization of bipartite graphs of readability 2.

6.12. Nonribosomal peptides

Norine is a comprehensive public database for non-ribosomal peptides developed by the team for more than 10 years. The Norine database quality has been enhanced through a semi-automatic curation process of data. Particularly, more than 500 SMILES annotations have been added or updated. This allowed us to check and correct the monomeric graphs, i.e. a 2D representation of the monomeric composition of the NRPs, thanks to dedicated tools like Smiles2Monomers. This update was done in collaboration with members of the Proteome Informatics Group from SIB (Swiss Institute of Bioinformatics). New annotations on monoisotopic mass and molecular formulas have also been added. The Norine interface was improved and new features are available, such as the possibility to access the complete change history of each entry. To encourage new submissions of NRPs, authors of new NRPs are now visible as contributors on Norine home page. Finally, we published this year, in the field of biocontrol (a contraction of "biological control"), a paper on bioinformatic tools for the discovery of new lipopeptides [5], essentially based on the Norine platform.

7. Partnerships and Cooperations

7.1. National Initiatives

7.1.1. ANR

- ANR Transipedia: The purpose of Transipedia is to provide means of identifying relevant transcriptomic events within thousands of RNA sequencing experiments. This project will be achieved in collaboration with I2BC (principal investigator) in Paris Saclay and IRMB in Montpellier.
- 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 work package devoted to the computational analysis of sRNA-seq data, in coordination with the bioinformatics platform of Génopole Toulouse-Midi-Pyrénées.

7.1.2. ADT

- ADT SeedLib (2017–2019): The SeedLib ADT aims to consolidate existing software developments in Bonsai, into an existing and well-engineered framework. Bonsai has published several new results on spaced seeds and developed several tools that integrate custom implementations of spaced seeds. In parallel, the GATB project is a C++ software library that facilitates the development of next-generation sequencing analysis tools. It is currently maintained by a collaboration between the GenScale team at Inria Rennes and the Bonsai team. Many users from other institutions (including the Erable team at Inria Rhones-Alpes) actively develop tools using GATB. The core object in GATB is k-mers, which can be seen as the predecessor of spaced seeds. The goal of this ADT is to integrate existing space seeds formalisms into GATB, therefore further expanding the features offered by the library, and at the same time provide visibility for tools and results in the Bonsai team.

7.2. European Initiatives

7.2.1. Collaborations in European Programs, Except FP7 & H2020

- International ANR RNAlands (2014-2018): 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.

7.3. International Research Visitors

7.3.1. Visits of International Scientists

7.3.1.1. Internships

- Inria MITACS 3-month internship of D. Martchenko (PhD student, Trent University)

8. Dissemination

8.1. Promoting Scientific Activities

8.1.1. Scientific Events Organisation

8.1.1.1. General Chair, Scientific Chair

- H. Touzet was the general chair for the satellites of RECOMB 2018 (Paris, April 2018)

8.1.2. Scientific Events Selection

8.1.2.1. Chair of Conference Program Committees

- RECOMB-Seq 2018 (R. Chikhi)

8.1.2.2. Member of the Conference Program Committees

- RECOMB-CG 2018 (J.-S. Varré)
- WABI 2018 (H. Touzet, M. Salson)
- ACM-BCB 2018 (R. Chikhi)
- HiCOMB 2018 (R. Chikhi)

8.1.2.3. Reviewer

- WABI 2018 (L. Noé, J.-S. Varré, R. Chikhi)
- ECCB 2018 (R. Chikhi)

8.1.3. Journal

8.1.3.1. Member of the Editorial Boards

- Review Editor in Frontiers in Genetics - Bioinformatics and Computational Biology (J.-S. Varré)

8.1.3.2. Reviewer - Reviewing Activities

- Algorithms (L. Noé)
- Bioinformatics (L. Noé, M. Salson, R. Chikhi)
- Frontiers in Immunology (M. Salson)
- Nucleic Acids Research (R. Chikhi)
- GigaScience (R. Chikhi)
- ACM Computing Surveys (R. Chikhi)
- Genome Biology (R. Chikhi)
- Journal of Discrete Algorithms (R. Chikhi)

8.1.4. Scientific Expertise

- Scientific consulting for Clarity Genomics start-up – Belgium (R. Chikhi)

8.1.5. Research Administration

- 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)
- Co-head of the Lille Bioinformatics core facility, bilille (H. Touzet)
- Member of the CRIStAL Laboratory council (H. Touzet)

8.2. Teaching - Supervision - Juries

8.2.1. Teaching

Teaching in computer science:

License: J.-S. Varré, *Programming and algorithms*, 36h, L2 Computer Science, Univ. Lille.

License: J.-S. Varré, *Object oriented programming*, 36h, L2 Computer Science, Univ. Lille.

License: J.-S. Varré, *Algorithms and data structures*, 50h, L2 Computer science, Univ. Lille.

License: J.-S. Varré, *System*, 84h, L3 Computer science, Univ. Lille.

License: P. Marijon *Databases*, 36h, L3 Computer science, Univ. Lille.

License: Q. Bonenfant *Databases*, 36h, L3 Computer science, Univ. Lille.

Licence: M. Pupin, *Programming (Python)*, 78h, L1 Sciences, Univ. Lille.

Licence: M. Pupin, *occupational integration*, 30h, L3 computer science, Univ. Lille.

Master: : M. Pupin, *Programming (JAVA)*, 24h, M1 "Mathématiques et finance", Univ. Lille.

License: M. Salson, *Programming (Python)*, 42h, L1 Sciences, Univ. Lille.

License: M. Salson, *Coding and information theory*, 63h, L2 Computer science, Univ. Lille.

Master: M. Salson, *Software project*, 40h, M1 Computer science, Univ. Lille.

Teaching in bioinformatics:

- Master: M. Pupin, M. Salson *Bioinformatics*, 34h, M1 "Biologie-Santé", Univ. Lille.
- Master: M. Salson, *Algorithms for life sciences*, 20h, M2 Complex models, algorithms and data, Univ. Lille.
- Master: R. Chikhi, *Bioinformatics*, 20h, M1 Computer Science, Univ. Lille.

Teaching in skeptical thinking:

- Master: M. Salson, *Skeptical thinking*, 27h, M2 Journalist and Scientist, ESJ, Univ. Lille.

Formation for academics:

- Bilille permanent training: C. Saad (*Variants*, 13h), R. Chikhi (*De novo assembly* and *Metagenomics de novo assembly*, 8h), H. Touzet (*DNA analyis* and *Metagenomics*, 8h), M. Salson (*RNA-seq analysis*, 1h).
- (JC)2BIM summer school: H. Touzet organized a national summer school in bioinformatics (5-8 June, Frejus), that gathered 30 participants and 12 trainers (among them, R. Chikhi and M. Salson): http://www.gdr-bim.cnrs.fr/?page_id=560.
- Workshop on Genomics: R. Chikhi (*de novo assembly & k-mers*, 8h), Czech Republic, 2 weeks, 80 participants: http://evomics.org/workshops/workshop-on-genomics/2018-workshop-on-genomics-cesky-krumlov/.
- CGSI summer school: R. Chikhi (*de novo assembly, metagenomics*, two keynote lectures), Los Angeles, 4 weeks, 100 participants: http://computationalgenomics.bioinformatics.ucla.edu.

8.2.2. Teaching administration

Head of the licence semester "Computer Science – S3 Harmonisation (S3H)", Univ. Lille (L. Noé).

Member of faculty council (M. Pupin, J.-S. Varré).

Head of the 3rd year of licence of computer science, Univ. Lille (J.-S. Varré).

Head of the GIS department (Software Engineering and Statistics) of Polytech'Lille (S. Janot).

Head of the computer science modules in the 1st year of Licence, Univ. Lille (M. Pupin).

Head of the *Informatique au féminin*, Univ. Lille (M. Pupin).

8.2.3. Supervision

PhD: T. Rocher, Indexing VDJ recombinations in lymphocytes for leukemia follow-up, February 2018, M. Giraud, M. Salson.

PhD: C. Saad, Caractérisation des erreurs de séquençage non aléatoires, application aux mosaïques et tumeurs hétérogènes, September 2018, M.-P. Buisine, H. Touzet, J. Leclerc, L. Noé, M. Figeac.

PhD in progress: P. Marijon, Analyse de graphes d'assemblage issus du séquençage ADN troisième génération, 2016, R. Chikhi, J-S. Varré.

PhD in progress: Q. Bonenfant, Algorithmes pour l'analyse de séquençage ARN troisième génération, 2017/11/15, L. Noé, H. Touzet.

8.2.4. Juries

- H. Touzet was member of the PhD juries of Camille Marchet (University of Rennes 1), Magali Dancette (University of Lyon 1), Aurélien Quillet (University of Rouen).
- H. Touzet was member of hiring committees (professors) at University of Lille and University of Caen.
- J.-S. Varré was member of a hiring committee (professors) at University of Lille.
- L. Noé was member of a hiring committee (associate professor) at University of Lille.
- R. Chikhi was member of the PhD juries of Florian Plaza Onate (INRA MetaGenoPolis), Sorina Maciuca (Wellcome Trust, UK).
- R. Chikhi was member of a hiring committee (research engineer) at Institut Pasteur, Paris.

8.3. Popularization

8.3.1. Internal or external Inria responsibilities

- Member of the CDT for Inria Lille (M. Pupin).
- Member of an Inria hiring committee (young researcher), Inria LNE (M. Pupin)

8.3.2. Education

- M. Pupin is the new leader of the collective *Informatique au féminin* from University of Lille, which was launched four years ago and whose goal is to organise computer science initiatives that reach teenage girls and female students. Among other actions, she is fully involved in the event *L codent, L créent* (she codes, she creates). This action aims to teach code to schoolgirls (13-15 years old), before they amass prejudices against computer science. 50 teenage girls were supervised by 11 female graduate computer science students, to create a proximity link between the young women. To emphasize the fact that coding is a creative and innovative pursuit, we chose to teach *Processing*, a programming language built for visual arts. After eight sessions of creative coding, a public exhibition was organized at the University with inspirational testimonies of women working in the field of computer science.
- 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", visit of high school students in the team (M. Salson).

9. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] T. ROCHER.*Compressing and indexing labeled sequences*, Université de Lille, February 2018, https://tel.archives-ouvertes.fr/tel-01758361

[2] C. SAAD.*Characterization of non-random sequencing errors, application to mosaicism and heterogeneous tumors*, Université de Lille Nord de France, September 2018, https://tel.archives-ouvertes.fr/tel-01936291

Articles in International Peer-Reviewed Journal

[3] V. G. CRAWFORD, A. KUHNLE, C. BOUCHER, R. CHIKHI, T. GAGIE.*Practical dynamic de Bruijn graphs*, in "Bioinformatics", June 2018 [*DOI :* 10.1093/BIOINFORMATICS/BTY500], https://hal.archives-ouvertes.fr/hal-01935559

[4] 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. DE 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. DE RIDDER, J. DE 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", 2018, vol. 19, n^{o} 1, p. 118-135 [*DOI :* 10.1093/BIB/BBW089], https://hal.inria.fr/hal-01390478

[5] M. PUPIN, A. FLISSI, P. JACQUES, V. LECLÈRE.*Bioinformatics tools for the discovery of new lipopeptides with biocontrol applications*, in "European Journal of Plant Pathology", July 2018 [*DOI :* 10.1007/S10658-018-1544-2], https://hal.archives-ouvertes.fr/hal-01937890

[6] T. ROCHER, M. GIRAUD, M. SALSON.*Indexing labeled sequences*, in "PeerJ Computer Science", 2018, vol. 4, p. 1-14 [*DOI :* 10.7717/PEERJ-CS.148], https://hal.archives-ouvertes.fr/hal-01743104

[7] C. SAAD, L. NOÉ, H. RICHARD, J. LECLERC, M.-P. BUISINE, H. TOUZET, M. FIGEAC.*DiNAMO: highly sensitive DNA motif discovery in high-throughput sequencing data*, in "BMC Bioinformatics", December 2018, vol. 19, n^{o} 1 [*DOI :* 10.1186/S12859-018-2215-1], https://hal.inria.fr/hal-01881466

Conferences without Proceedings

[8] R. CHIKHI, V. JOVIČIĆ, S. KRATSCH, P. MEDVEDEV, M. MILANIC, S. RASKHODNIKOVA, N. VARMA.*Bipartite Graphs of Small Readability*, in "COCOON 2018 - The 24th International Computing and Combinatorics Conference", Qingdao, China, July 2018, https://arxiv.org/abs/1805.04765 , https://hal.archives-ouvertes.fr/hal-01935562

[9] A. KUOSMANEN, T. PAAVILAINEN, T. GAGIE, R. CHIKHI, A. I. TOMESCU, V. MAKINEN.*Using Minimum Path Cover to Boost Dynamic Programming on DAGs: Co-Linear Chaining Extended*, in "RECOMB 2018 - 22nd Annual International Conference on Research in Computational Molecular Biology", Paris, France, April 2018, https://arxiv.org/abs/1705.08754 , https://hal.archives-ouvertes.fr/hal-01935568

[10] P. MARQUET, M. PUPIN, Y. SECQ.*L codent, L créent: créations numériques artistiques pour démystifier l'informatique... au féminin! (descriptif d'atelier)*, in "Didapro 7 – DidaSTIC. De 0 à 1 ou l'heure de l'informatique à l'école", Lausanne, Switzerland, February 2018, p. 1-2, https://hal.archives-ouvertes.fr/hal-01753402

Scientific Popularization

[11] R. DAVID, L. MABILE, M. YAHIA, A. CAMBON-THOMSEN, A.-S. ARCHAMBEAU, L. BEZUIDENHOUT, S. BEKAERT, G. BERTIER, E. BRAVO, J. CARPENTER, A. COHEN-NABEIRO, A. DELAVAUD, M. DE ROSA, L. DOLLÉ, F. GRATTAROLA, F. MURPHY, S. PAMERLON, A. SPECHT, A.-M. TASSÉ, M. THOMSEN, M. ZILIOLI.*Operationalizing and evaluating the FAIRness concept for a good quality of data sharing in Research: the RDA-SHARC-IG (SHAring Rewards and Credit Interest Group)*, November 2018, assemblée MaDICS 2018, Poster [*DOI :* 10.5281/ZENODO.1745374], https://hal.archives-ouvertes.fr/hal-01929834

Other Publications

[12] R. CHIKHI, A. SCHÖNHUTH.*Dualities in Tree Representations*, November 2018, https://arxiv.org/abs/1804.04263 - CPM 2018, extended version, https://hal.archives-ouvertes.fr/hal-01935566

Team BONUS

Big Optimization aNd Ultra-Scale computing

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

1. Team, Visitors, External Collaborators **25**
2. Overall Objectives **26**
3. Research Program **27**
3.1. Decomposition-based Optimization 27
3.2. Machine Learning-assisted Optimization 29
3.3. Ultra-scale Optimization 30
4. Application Domains **31**
4.1. Introduction 31
4.2. Big optimization for complex scheduling 31
4.3. Big optimization for engineering design 32
5. Highlights of the Year **32**
6. New Software and Platforms **32**
7. New Results **33**
7.1. Decomposition-based optimization 33
7.2. ML-assisted optimization 34
7.3. Large scale GPU-centric optimization 36
8. Bilateral Contracts and Grants with Industry **37**
9. Partnerships and Cooperations **37**
9.1. Regional Initiatives 37
9.2. National Initiatives 37
9.3. European Initiatives 38
9.3.1. FP7 & H2020 Projects 38
9.3.2. Collaborations in European Programs, Except FP7 & H2020 38
9.3.3. Collaborations with Major European Organizations 38
9.4. International Initiatives 38
9.4.1. Inria International Labs 38
9.4.1.1. MOHA 39
9.4.1.2. Other IL projects 39
9.4.2. Inria Associate Teams Not Involved in an Inria International Labs 39
9.4.3. Inria International Partners 39
9.4.4. Participation in Other International Programs 39
9.5. International Research Visitors 40
10. Dissemination **40**
10.1. Promoting Scientific Activities 40
10.1.1. Scientific Events Organisation 40
10.1.1.1. General Chair, Scientific Chair 40
10.1.1.2. Member of the Organizing Committees 41
10.1.2. Scientific Events Selection 41
10.1.2.1. Chair of Conference Program Committees 41
10.1.2.2. Member of the Conference Program Committees 41
10.1.3. Journal 41
10.1.3.1. Member of the Editorial Boards 41
10.1.3.2. Reviewer - Reviewing Activities 42
10.1.4. Invited Talks 42
10.1.5. Leadership within the Scientific Community 42
10.1.6. Scientific Expertise 42
10.1.7. Research Administration 43
10.2. Teaching - Supervision - Juries 43
10.2.1. Teaching 43

10.2.2. Supervision 44
10.2.3. Juries 44
10.3. Popularization 44
11. Bibliography **44**

Team BONUS

Creation of the Team: 2017 July 01

Keywords:

Computer Science and Digital Science:

A6.2.7. - High performance computing
A8.2.1. - Operations research
A8.2.2. - Evolutionary algorithms
A9.6. - Decision support
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B2. - Health
B4. - Energy
B7. - Transport and logistics

1. Team, Visitors, External Collaborators

Faculty Members

Nouredine Melab [Team leader, Univ des sciences et technologies de Lille, Professor, HDR]
Omar Abdelkafi [Univ des sciences et technologies de Lille, Associate Professor]
Bilel Derbel [Deputy leader, Univ des sciences et technologies de Lille, Associate Professor, HDR]
Arnaud Liefooghe [Univ des sciences et technologies de Lille, Associate Professor]
El-Ghazali Talbi [Univ des sciences et technologies de Lille, Professor, HDR]

Technical Staff

Jingyu Ji [Inria, from Feb 2018]

PhD Students

Nicolas Berveglieri [Univ des sciences et technologies de Lille, from Oct 2018]
Guillaume Briffoteaux [Univ des sciences et technologies de Lille]
Sohrab Faramarzi Oghani [Inria]
Maxime Gobert [Univ des sciences et technologies de Lille]
Ali Hebbal [ONERA]
Julien Pelamatti [ONERA]
Geoffrey Pruvost [Univ des sciences et technologies de Lille, from Oct 2018]
Jeremy Sadet [Univ de Valenciennes et du Hainaut Cambrésis, from Oct 2018]

Post-Doctoral Fellows

Tiago Carneiro Pessoa [Inria, from Nov 2018]
Mohammad Rahimi [Univ des sciences et technologies de Lille]
Jan Gmys [Université de Mons, ATER]
Oumayma Bahri [Univ des sciences et technologies de Lille, ATER]

Visiting Scientists

Alexandre Jesus [University of Coimbra, Portugal, January 2018]
Kiyoshi Tanaka [Shinshu University, Japan, March 2018 and November 2018]
Kalyan Deb [University of Michiga, USA, Oct 2018]
Rachid Ellaia [EMI Univ. Rabat, Morocco, April 2018]
Hernan Aguirre [Shinshu University, Japan, Invited Professor Univ Lille, from March 2018 until April 2018]
Mariem Belhor [ENSI, Tunis, TUNISIA, from Apr 2018 until May 2018]

Administrative Assistants

Julie Jonas [Inria, until June 2018]

Karine Lewandowski [Inria, from June 2018]

2. Overall Objectives

2.1. Presentation

In the BONUS project, the context of optimization, where solving a problem consists in optimizing (minimizing or maximizing) one or more objective function(s) under some constraints, is considered. In this context, a problem can be formulated as follows:

$$\begin{aligned} &\text{Min/Max } F(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), ..., f_m(\mathbf{x})) \\ &\text{subject to} \qquad \mathbf{x} \in S. \end{aligned}$$

where S is the feasible search space and $\mathbf{x}$ is the decision variable vector of dimension n.

Nowadays, in many research and application areas we are witnessing the emergence of the big era (big data, big graphs, etc). In the optimization setting, the problems are increasingly big in practice. Big optimization problems (BOPs) refer to problems composed of a large number of environmental input parameters and/or decision variables (*high dimensionality*), and/or *many objective functions* that may be *computationally expensive*. For instance, in smart grids, there are many optimization problems for which have to be considered a large number of consumers (appliances, electrical vehicles, etc.) and multiple suppliers with various energy sources. In the area of engineering design, the optimization process must often take into account a large number of parameters from different disciplines. In addition, the evaluation of the objective function(s) often consist(s) in the execution of an expensive simulation of a black-box complex system. This is for instance typically the case in aerodynamics where a CFD-based simulation may require several hours. On the other hand, to meet the high growing needs of applications in terms of computational power in a wide range of areas including optimization, high-performance computing (HPC) technologies have known a revolution during the last decade (see Top500 [0]). Indeed, HPC is evolving toward *ultra-scale supercomputers composed of millions of cores supplied in heterogeneous devices including multi-core processors with various architectures, GPU accelerators and MIC coprocessors*.

Beyond the "big buzzword", solving BOPs raises at least four major challenges: (1) tackling their high dimensionality; (2) handling many objectives; (3) dealing with computationally expensive objective functions; and (4) scaling on (ultra-scale) modern supercomputers. The overall scientific objectives of the BONUS project consist in addressing efficiently these challenges. On the one hand, the focus will be put on the design, analysis and implementation of optimization algorithms scalable to high-dimensional (in decision variables and/or objectives) and/or expensive problems. On the other hand, the focus will also be put on the design of optimization algorithms able to scale on heterogeneous supercomputers including several millions of processing cores. To achieve these objectives raising the associated challenges a program including three lines of research will be adopted (Fig. 1): *decomposition-based optimization, Machine Learning (ML)-assisted optimization and ultra-scale optimization*. These research lines are developed in the following section.

From the software standpoint, our objective is to integrate the approaches we will develop in our *ParadisEO* [3] [0] framework in order to allow their reuse inside and outside the BONUS team. The major challenge will be to extend ParadisEO in order to make it *more collaborative* with other software including machine learning tools, other (exact) solvers and simulators. *From application point of view*, the focus will be put on two classes of applications: *complex scheduling and engineering design*.

[0]Top500 international ranking (Edition of November 2018): https://www.top500.org/lists/2018/11/

[0]ParadisEO: http://paradiseo.gforge.inria.fr/

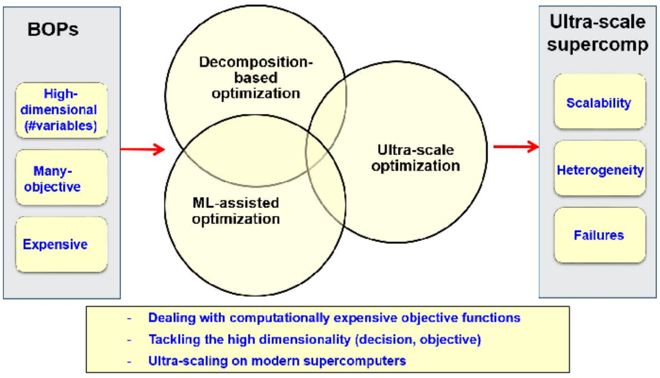

Figure 1. Research challenges/objectives and lines

3. Research Program

3.1. Decomposition-based Optimization

Given the large scale of the targeted optimization problems in terms of the number of variables and objectives, their decomposition into simplified and loosely coupled or independent subproblems is essential to raise the challenge of scalability. The first line of research is to *investigate the decomposition approach in the two spaces and their combination, as well as their implementation on ultra-scale architectures*. The motivation of the decomposition is twofold: first, the decomposition allows the parallel resolution of the resulting subproblems on ultra-scale architectures. Here also several issues will be addressed: the definition of the subproblems, their coding to allow their efficient communication and storage (checkpointing), their assignment to processing cores, etc. Second, decomposition is necessary for solving large problems that cannot be solved (efficiently) using traditional algorithms. Indeed, for instance with the popular NSGA-II algorithm the number of non-dominated solutions [0] increases drastically with the number of objectives leading to a very slow convergence to the Pareto front [0]. Therefore, decomposition-based techniques are gaining a growing interest. The objective of BONUS is to *investigate various decomposition schema and cooperation protocols between the subproblems* resulting from the decomposition to generate efficiently global solutions of good quality. Several challenges have to be addressed: how to define the subproblems (decomposition strategy), how to solve them to generate local solutions (local rules), and how to combine these latter with those generated by other subproblems and how to generate global solutions (cooperation mechanism), and how to combine decomposition strategies in more than one space (hybridization strategy)? These challenges, which are in the line with the CIS Task Force [0] on decomposition will be addressed in the decision as well as in the objective space.

[0] A solution x dominates another solution y if x is better than y for all objectives and there exists at least one objective for which x is strictly better than y.

[0] A Pareto Front is the set of non-dominated solutions.

[0] IEEE CIS Task Force, created in 2017 on Decomposition-based Techniques in Evolutionary Computation.

The *decomposition in the decision space* can be performed following different ways according to the problem at hand. Two major categories of decomposition techniques can be distinguished: the first one consists in *breaking down the high-dimensional decision vector* into low-dimensional and easier-to-optimize blocks of variables. The major issue is how to define the subproblems (blocks of variables) and their cooperation protocol: randomly *vs.* using some learning (e.g. separability analysis), statically *vs.* adaptively, etc. *The decomposition in the decision space can also be guided by the type of variables i.e. discrete vs. continuous*. The discrete and continuous parts are optimized separately using cooperative hybrid algorithms [47]. *The major issue of this kind of decomposition is the presence of categorial variables in the discrete part [33]. The* BONUS *team is addressing this issue, rarely investigated in the literature*, within the context of vehicle aerospace engineering design. The second category consists in the *decomposition according to the ranges of the decision variables*. For continuous problems, the idea consists in iteratively subdividing the search (e.g. design) space into subspaces (hyper-rectangles, intervals, etc.) and select those that are most likely to produce the lowest objective function value. *Existing approaches meet increasing difficulty with an increasing number of variables and are often applied to low-dimensional problems. We are investigating this scalability challenge* (e.g. [10]). *For discrete problems, the major challenge is to find a coding (mapping) of the search space to a decomposable entity*. We have proposed an interval-based coding of the permutation space for solving big permutation problems. The approach opens perspectives we are investigating [7], in terms of ultra-scale parallelization, application to multi-permutation problems and hybridization with metaheuristics.

The *decomposition in the objective space* consists in breaking down an original MaOP into a set of cooperative single-objective subproblems (SOPs). The decomposition strategy requires the careful definition of a scalarizing (aggregation) function and its weighting vectors (each of them corresponds to a separate SOP) to guide the search process towards the best regions. Several scalarizing functions have been proposed in the literature including weighted sum, weighted Tchebycheff, vector angle distance scaling, etc. These functions are widely used but they have their limitations. For instance, using weighted Tchebycheff might do harm diversity maintenance and weighted sum is inefficient when it comes to deal with nonconvex Pareto Fronts [40]. Defining a scalarizing function well-suited to the MaOP at hand is therefore a difficult and still an open question being investigated in BONUS [6], [5]. Studying/defining various functions and in-depth analyzing them to better understand the differences between them is required. Regarding the weighting vectors that determine the search direction, their efficient setting is also a key and open issue. They dramatically affect in particular the diversity performance. Their setting rises several issues mainly: how to determine their number according to the available computational ressources? when (statically or adaptively) and how to determine their values? *Weight adaptation is one of our main concerns that we are addressing especially from a distributed perspective.* They correspond to the main scientific objectives targeted by our bilateral ANR-RGC BigMO project with City University (Hong Kong). The other challenges pointed out in the beginning of this section concern the way to solve locally the SOPs resulting from the decomposition of a MaOP and the mechanism used for their cooperation to generate global solutions. To deal with these challenges, our approach is to design the decomposition strategy and cooperation mechanism keeping in mind the parallel and/or distributed solving of the SOPs. Indeed, we favor the local neighborhood-based mating selection and replacement to minimize the network communication cost while allowing an effective resolution [5]. The major issues here are how to define the neighborhood of a subproblem and how to cooperatively update the best-known solution of each subproblem and its neighbors.

To sum up, the objective of the BONUS *team is to come up with scalable decomposition-based approaches in the decision and objective spaces. In the decision space, a particular focus will be put on high dimensionality and mixed-continuous variables which have received little interest in the literature. We will particularly continue to investigate at larger scales using ultra-scale computing the interval-based (discrete) and fractal-based (continuous) approaches. We will also deal with the rarely addressed challenge of mixed-continuous including categorial variables (collaboration with ONERA). In the objective space, we will investigate parallel ultra-scale decomposition-based many-objective optimization with ML-based adaptive building of scalarizing functions. A particular focus will be put on the state-of-the-art MOEA/D algorithm. This challenge is rarely addressed in the literature which motivated the collaboration with the designer of MOEA/D (bilateral ANR-*

RGC BigMO project with City University, Hong Kong). Finally, the joint decision-objective decomposition, which is still in its infancy [49], is another challenge of major interest.

3.2. Machine Learning-assisted Optimization

The Machine Learning (ML) approach based on metamodels (or surrogates) is commonly used, and also adopted in BONUS, to assist optimization in tackling BOPs characterized by time-demanding objective functions. The second line of research of BONUS is focused on ML-aided optimization to raise the challenge of expensive functions of BOPs using surrogates but also to assist the two other research lines (decomposition-based and ultra-scale optimization) in dealing with the other challenges (high dimensionality and scalability).

Several issues have been identified to make efficient and effective surrogate-assisted optimization. First, infill criteria have to be carefully defined to adaptively select the adequate sample points (in terms of surrogate precision and solution quality). The challenge is to find the best trade-off between exploration and exploitation to efficiently refine the surrogate and guide the optimization process toward the best solutions. The most popular infill criterion is probably the *Expected Improvement* (EI) [43] which is based on the expected values of sample points but also and importantly on their variance. This latter is inherently determined in the kriging model, this is why it is used in the state-of-the-art *efficient global optimization* (EGO) algorithm [43]. However, such crucial information is not provided in all surrogate models (e.g. ANN) and needs to be derived. In BONUS, we are currently investigating this issue. Second, it is known that surrogates allow one to reduce the computational burden for solving BOPs with time-demanding function(s). However, using parallel computing as a complementary way is often recommended and cited as a perspective in the conclusions of related publications. Nevertheless, *despite being of critical importance parallel surrogate-assisted optimization is weakly addressed in the literature*. For instance, in the introduction of the survey proposed in [42] it is warned that because the area is not mature yet the paper is more focused on the potential of the surveyed approaches than on their relative efficiency. *Parallel computing is required at different levels that we are investigating*.

Another issue with surrogate-assisted optimization is related to high dimensionality in decision as well as in objective space: it is often applied to low-dimensional problems. *The joint use of decomposition, surrogates and massive parallelism is an efficient approach to deal with high dimensionality. This approach adopted in* BONUS *has received little effort in the literature.* In BONUS, we are considering a generic framework in order to enable a flexible coupling of existing surrogate models within the state-of-the-art decomposition-based algorithm MOEA/D. This is a first step in leveraging the applicability of efficient global optimization into the multi-objective setting through parallel decomposition. Another issue which is a consequence of high dimensionality is the mixed (discrete-continuous) nature of decision variables which is frequent in real-world applications (e.g. engineering design). *While surrogate-assisted optimization is widely applied in the continuous setting it is rarely addressed in the literature in the discrete-continuous framework.* In [33], we have identified different ways to deal with this issue that we are investigating. Non-stationary functions frequent in real-world applications (see Section 4.1) is another major issue we are addressing using the concept of deep GP.

Finally, as quoted in the beginning of this section, ML-assisted optimization is mainly used to deal with BOPs with expensive functions but it will also be investigated for other optimization tasks. Indeed, ML will be useful to assist the decomposition process. In the decision space, it will help to perform the separability analysis (understanding of the interactions between variables) to decompose the vector of variables. In the objective space, ML will be useful to assist a decomposition-based many-objective algorithm in dynamically selecting a scalarizing function or updating the weighting vectors according to their performances in the previous steps of the optimization process [5]. Such a data-driven ML methodology would allow us to understand what makes a problem difficult or an optimization approach efficient, to predict the algorithm performance [4], to select the most appropriate algorithm configuration [8], and to adapt and improve the algorithm design for unknown optimization domains and instances. Such an autonomous optimization approach would adaptively adjust its internal mechanisms in order to tackle cross-domain BOPs.

In a nutshell, to deal with expensive optimization the BONUS *team will investigate the surrogate-based ML approach with the objective to efficiently integrate surrogates in the optimization process. The focus will especially be put on high dimensionality (e.g. using decomposition) with mixed discrete-continuous variables which is rarely investigated. The kriging metamodel (Gaussian Process or GP) will be considered in particular for engineering design (for more reliability) addressing the above issues and other major ones including mainly non stationarity (using emerging deep GP) and ultra-scale parallelization (highly needed by the community). Indeed, a lot of work has been reported on deep neural networks (deep learning) surrogates but not on the others including (Deep) GP. On the other hand, ML will be used to assist decomposition: importance/interaction between variables in the decision space, dynamic building (selection of scalarizing functions, weight update, ...) of scalarizing functions in the objective space, etc.*

3.3. Ultra-scale Optimization

The third line of our research program that accentuates our difference from other (project-)teams of the related Inria scientific theme is the ultra-scale optimization. *This research line is complementary to the two others, which are sources of massive parallelism* and with which it should be combined to solve BOPs. Indeed, ultra-scale computing is necessary for the effective resolution of the large amount of subproblems generated by decomposition of BOPs, parallel evaluation of simulation-based fitness and metamodels, etc. These sources of parallelism are attractive for solving BOPs and are natural candidates for ultra-scale supercomputers [0]. However, their efficient use raises a big challenge consisting in managing efficiently a massive amount of irregular tasks on supercomputers with multiple levels of parallelism and heterogeneous computing resources (GPU, multi-core CPU with various architectures) and networks. Raising such challenge requires to tackle three major issues, scalability, heterogeneity and fault-tolerance, discussed in the following.

The *scalability* issue requires, on the one hand, the definition of scalable data structures for efficient storage and management of the tremendous amount of subproblems generated by decomposition [45]. On the other hand, achieving extreme scalability requires also the optimization of communications (in number of messages, their size and scope) especially at the inter-node level. For that, we target the design of asynchronous locality-aware algorithms as we did in [41], [48]. In addition, efficient mechanisms are needed for granularity management and coding of the work units stored and communicated during the resolution process.

Heterogeneity means harnessing various resources including multi-core processors within different architectures and GPU devices. The challenge is therefore to design and implement hybrid optimization algorithms taking into account the difference in computational power between the various resources as well as the resource-specific issues. On the one hand, to deal with the heterogeneity in terms of computational power, we adopt in BONUS the dynamic load balancing approach based on the Work Stealing (WS) asynchronous paradigm [0] at the inter-node as well as at the intra-node level. We have already investigated such approach, with various victim selection and work sharing strategies in [48], [7]. On the other hand, hardware resource specific-level optimization mechanisms are required to deal with related issues such as thread divergence and memory optimization on GPU, data sharing and synchronization, cache locality, and vectorization on multi-core processors, etc. These issues have been considered separately in the literature including our works [9], [1]. Indeed, in most of existing works related to GPU-accelerated optimization only a single CPU core is used. This leads to a huge resource wasting especially with the increase of the number of processing cores integrated into modern processors. Using jointly the two components raises additional issues including data and work partitioning, the optimization of CPU-GPU data transfers, etc.

Another issue the scalability induces is the *increasing probability of failures* in modern supercomputers [46]. Indeed, with the increase of their size to millions of processing cores their MTBF tends to be shorter and shorter [44]. Failures may have different sources including hardware and software faults, silent errors, etc.

[0] In the context of BONUS, supercomputers are composed of several massively parallel processing nodes (inter-node parallelism) including multi-core processors and GPUs (intra-node parallelism).

[0] A WS mechanism is mainly defined by two components: a victim selection strategy which selects the processing core to be stolen and a work sharing policy which determines the part and amount of the work unit to be given to the thief upon WS request.

In our context, we consider failures leading to the loss of work unit(s) being processed by some thread(s) during the resolution process. The major issue, which is particularly critical in exact optimization, is how to recover the failed work units to ensure a reliable execution. Such issue is tackled in the literature using different approaches: algorithm-based fault tolerance, checkpoint/restart (CR), message logging and redundancy. The CR approach can be system-level, library/user-level or application-level. Thanks to its efficiency in terms of memory footprint, adopted in BONUS [2], the application-level approach is commonly and widely used in the literature. This approach raises several issues mainly: what is critical information which defines the state of the work units and allows to resume properly their execution? when, where and how (using which data structures) to store it efficiently? how to deal with the two other issues: scalability and heterogeneity?

The last but not least major issue which is another roadblock to exascale is the programming of massive-scale applications for modern supercomputers. *On the path to exascale, we will investigate the programming environments and execution supports able to deal with exascale challenges: large numbers of threads, heterogeneous resources, etc.* Various exascale programming approaches are being investigated by the parallel computing community and HPC builders: extending existing programming languages (e.g. DSL-C++) and environments/libraries (MPI+X, etc.), proposing new solutions including mainly PGAS-based environments (Chapel, UPC, X10, etc.). It is worth noting here that our objective is not to develop a programming environment nor a runtime support for exascale computing. Instead, we aim to collaborate with the research teams (inside or outside Inria) having such objective.

To sum up, we put the focus on the design and implementation of efficient big optimization algorithms dealing jointly (uncommon in parallel optimization) with the major issues of ultra-scale computing mainly the scalability up to millions of cores using scalable data structures and asynchronous locality-aware work stealing, heterogeneity addressing the multi-core and GPU-specific issues and those related to their combination, and scalable GPU-aware fault tolerance. A strong effort will be devoted to this latter challenge, for the first time to the best of our knowledge, using application-level checkpoint/restart approach to deal with failures.

4. Application Domains

4.1. Introduction

For the validation of our findings we obviously use standard benchmarks to facilitate the comparison with related works. In addition, we also target real-world applications in the context of our collaborations and industrial contracts. From the *application* point of view two classes are targeted: *complex scheduling* and *engineering design*. The objective is twofold: proposing new models for complex problems and solving efficiently BOPs using jointly the three lines of our research program. In the following, are given some use cases that are the focus of our current industrial collaborations.

4.2. Big optimization for complex scheduling

Three application domains are targeted: energy, health and transport and logistics. In the **energy** field, with the smart grid revolution (multi-)house energy management is gaining a growing interest. The key challenge is to make elastic with respect to the energy market the (multi-)house energy consumption and management. *This kind of demand-side management will be of strategic importance for energy companies in the near future.* In collaboration with the EDF energy company we are working on the formulation and solving of optimization problems on demand-side management in smart micro-grids for single- and multi-user frameworks. These complex problems require taking into account multiple conflicting objectives and constraints and many (deterministic/uncertain, discrete/continuous) parameters. A representative example of such BOPs that we are addressing is the scheduling of the activation of a large number of electrical and thermal appliances for a set of homes optimizing at least three criteria: maximizing the user's confort, minimizing its energy bill and minimzing peak consumption situations. In the **health** care domain, we are collaborating with the Beckman &

Coulter company on the design and planning of large medical laboratories. This is a hot topic resulting from the mutualisation phenomenon which makes bigger these laboratories. As a consequence, being responsible for analyzing medical tests ordered by physicians on patient's samples, these laboratories receive large amounts of prescriptions and tubes making their associated workflow more complex. Our aim is therefore to design and plan any medical laboratory to minimize the costs and time required to perform the tests. More exactly, the focus is put on the multi-objective modeling and solving of large (e.g. dozens of thousands of medical test tubes to be analyzed) strategic, tactical and operational problems such as the layout design, machine selection and confguration, assignment and scheduling. Finally, in **transport and logistics**, within the context of our potential collaboration (being set up) with the EXOTEC company we target the optimization of the robotic logistics of 3D warehouses. More exactly, the problem consists in efficient complex scheduling without collision of thousands of missions realized by a fleet of dozens of robots and several operators in a 3D logistics warehouse. The problem is identified in the literature as the parts-to-picker based order processing in a rack-moving mobile robots environment.

4.3. Big optimization for engineering design

The focus is for now put on the aerospace vehicle design, a complex multidisciplinary optimization process, we are exploring in collaboration with ONERA. The objective is to find the vehicle architecture and characteristics that provide the optimal performance (flight performance, safety, reliability, cost, etc.) while satisfying design requirements [39]. A representative topic we are investigating, and will continue to investigate throughout the lifetime of the project given its complexity, is the design of launch vehicles that involves at least 4 tightly coupled disciplines (aerodynamics, structure, propulsion and trajectory). Each discipline may rely on time-demanding simulations such as Finite Element analyses (structure) and Computational Fluid Dynamics analyses (aerodynamics). Surrogate-assisted optimization is highly required to reduce the time complexity. In addition, the problem is high-dimensional (dozens of parameters and more than 3 objectives) requiring different decomposition schema (coupling *vs.* local variables, continuous *vs.* discrete even categorial variables, scalarization of the objectives). Another major issue arising in this area is the non-stationarity of the objective functions which is generally due to the abrupt change of a physical property that often occurs in the design of launch vehicles. In the same spirit than deep learning using neural networks, we use Deep Gaussian Processes to deal with non-stationary multi-objective functions. Finally, the resolution of the problem using only one objective takes 1 week using a multi-core processor. Therefore, *in addition to surrogates ultra-scale computing is required at different levels to speed up the search and improve the reliability which is a major requirement in aerospace design.* This example shows that we need to use the synergy between the 3 lines of our research program to tackle such BOPs.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Awards

- Patent with Beckman & Coulter on the optimization of large medical laboratories (Prof. E-G. Talbi, S. Faramarzi-oghani, M. Bué).

6. New Software and Platforms

6.1. Platforms

6.1.1. Grid'5000 testbed: extension with GPUs at Lille

KEYWORDS: Experimental testbed, large-scale computing, high-performance computing, GPU computing, cloud computing, big data

FUNCTIONAL DESCRIPTION: Grid'5000 is a project initiated in 2003 by the French government to promote scientific research on large scale distributed systems. The project is later supported different research organizations including Inria, CNRS, the french universities, Renater which provides the wide-area network, etc. The overall objective of Grid'5000 was to build by 2007 a nation-wide experimental testbed composed of at least 5000 processing units and distributed over several sites in France. From a scientific point of view, the aim was to promote scientific research on large-scale distributed systems.
Grid'5000 was installed at the center of IT resources including supercomputing resources of Université de Lille 1 and opened to users in 2005. Since March 2017, the Grid'5000 site has moved to the premises on Inria Lille within the context of the phase 1 of the CPER data program (see Section 9.1) with a completely new hardware equipment. As a scientific leader of the testbed for the Lille's site, N. Melab has been strongly involved in the extension (phase 2 of CPER data) of the platform with 16 computing serveurs, 16 Nvidia GPUs (12 P100 and 4V100), 2 storage serveurs 200TB and 2 administration servers. Grid'5000 at Lille is used by more than 150 users including 100 external ones. The testbed is used for research as well as for teaching allowing a high scientific production (publications, PhD theses, etc.) and over 30 master students to get started with parallel and distributed programming.

- Participants: N. Melab, external collaborators: D. Delabroy, T. Peltier, L. Nussbaum.
- Contact: Nouredine Melab.
- URL: https://www.grid5000.fr/mediawiki/index.php/Grid5000:Home

7. New Results

7.1. Decomposition-based optimization

- **A set-oriented decomposition algorithm for multi-objective optimization.**
  Participants: B. Derbel and A. Liefooghe, external collaborators: H. Aguirre and K. Tanaka, Shinshu University (JAPAN); S. Verel, Univ. Littoral (FRANCE); Q. Zhang, City University (HONG KONG)

  The working principles of the well-established multi-objective evolutionary algorithm MOEA/D relies on the iterative and cooperative improvement of a number of single-objective sub-problems obtained by decomposition. Besides the definition of sub-problems, selection and replacement are, like in any evolutionary algorithm, the two core elements of MOEA/D. We argue that these two components are however loosely coupled with the maintained population. Thereby, in [24], we propose to re-design the working principles of MOEA/D by adopting a set-oriented perspective, where a many-to-one mapping between sub-problems and solutions is considered. Selection is then performed by defining a neighborhood relation among solutions in the population set, depending on the corresponding sub-problem mapping. Replacement is performed following an elitist mechanism allowing the population to have a variable, but bounded, cardinality during the search process. By conducting a comprehensive empirical analysis on a range of combinatorial multi- and many-objective nk-landscapes, we show that the proposed approach leads to significant improvements, especially when dealing with an increasing number of objectives. Our findings indicate that a set-oriented design can constitute a sound alternative for strengthening the practice of multi- and many-objective evolutionary optimization based on decomposition.
- **Parallel Pareto local search for multi-objective optimization.**
  Participants: B. Derbel and A. Liefooghe, external collaborators: J. Shi and J. Sun, Xi'an Jiaotong University (CHINA); Q. Zhang, City University (HONG KONG)

  Pareto Local Search (PLS) is a simple, yet effective optimization approach dedicated to multi-objective combinatorial optimization. It can however suffer from a high computational cost, especially when the size of the Pareto optimal set is relatively large. Recently, incorporating decomposition in PLS had revealed a high potential, not only in providing high-quality approximation sets, but also in speeding-up the search process. In [30], using the bi-objective Unconstrained Binary Quadratic Programming (bUBQP) problem as an illustrative benchmark, we demonstrate some shortcomings in the resulting decomposition-guided Parallel Pareto Local Search (PPLS), and we propose to revisit the PPLS design accordingly. For instances with a priori unknown Pareto front shape, we show that a simple pre-processing technique to estimate the scale of the Pareto front can help PPLS to better balance the workload. Furthermore, we propose a simple technique to deal with the critically-important scalability issue raised by PPLS when deployed over a large number of computing nodes. Our investigations show that the revisited version of PPLS provides a consistent performance, suggesting that decomposition-guided PPLS can be further generalized in order to improve both parallel efficiency and approximation quality.

- **Archivers for the representation of the set of approximate solutions for MOPs.**
  Participants: E-G. Talbi, external collaborators: O. Schutze, C. Hernandez (Computer Science Department, Cinvestav, MEXICO), Q. Sun, Y. Naranjani (School of Engineering University of California, USA), R. Xiong (Department of Mechanics, University Tianjin, CHINA)

  In this work we have addressed the problem of computing suitable representations of the set of approximate solutions of a given multi-objective optimization problem via stochastic search algorithms. For this, we have proposed different archiving strategies for the selection of the candidate solutions maintained by the generation process of the stochastic search process, and investigate them further on analytically and empirically. For all archivers we have provided upper bounds on the approximation quality as well as on the cardinality of the limit solution set. A comparative study on some test problems in order to visualize the effect of all novel archiving strategies has also been carried out [18].

7.2. ML-assisted optimization

FIve major contributions related to ML-assisted optimization have been achieved and summarized in the following. As pointed out previously in our research program, one of the major issues in surrogate-assisted optimization is how to integrate efficiently and effectively the surrogates in the optimization process. This issue is addressed in first three contributions. Another major aspect addressed in the fourth contribution is the investigation of surrogates within the context of combinatorial optimization. The focus of the fifth contribution is put on the landscape analysis applied within the context of multi-objective optimization.

- **Efficient Global Optimization Using Deep Gaussian Processes.**
  Participants: A. Hebbal, E-G. Talbi and N. Melab, external collaborators: L. Brevault and M. Balesdent from ONERA (Palaiseau, Paris)

  Efficient Global Optimization (EGO) is widely used for the optimization of computationally expensive black-box functions. EGO is based on a surrogate modeling technique using Gaussian Processes (kriging). However, due to the use of a stationary covariance, kriging is not well suited for approximating non stationary functions. Non stationarity is generally due to the abrupt change of a physical property that often occurs in the design of launch vehicles, subject of our collaboration with ONERA. This leads to a variation of the objective function with a completely different smoothness along the input space. In the spirit of deep learning using neural networks, we have investigated in [25] the integration of Deep Gaussian processes (DGP) in EGO framework to deal with non stationarity. Numerical experimentations are performed on analytical problems to highlight the different aspects of DGP and EGO. The experimental results show that the coupling EGO-DGP outperforms EGO-GP with a significant margin. Furthermore, the study has also highlighted some challenging issues to be investigated including: the integration of DGP in multi-objective EGO, the configuration of the network and revisiting the training model. Ultra-scale optimization at different levels is particularly important given the large number of hyperparameters of the training model.

- **Efficient global optimization of constrained mixed variable problems.**
  Participants: E-G. Talbi, external collaborators: Julien Pelamatti, Loïc Brevault, Mathieu Balesdent (ONERA) Yannick Guerin (CNES)

Due to the increasing demand for high performance and cost reduction within the framework of complex system design, numerical optimization of computationally costly problems is an increasingly popular topic in most engineering fields [33]. In this work, several variants of the Efficient Global Optimization algorithm for costly constrained problems depending simultaneously on continuous decision variables as well as on quantitative and/or qualitative discrete design parameters are proposed. The adaptation that is considered is based on a redefinition of the Gaussian Process kernel as a product between the standard continuous kernel and a second kernel representing the covariance between the discrete variable values. Several parameterizations of this discrete kernel, with their respective strengths and weaknesses, have been investigated. The novel algorithms are tested on a number of analytical test-cases and an aerospace related design problem, and it is shown that they require fewer function evaluations in order to converge towards the neighborhoods of the problem optima when compared to more commonly used optimization algorithms [38].

- **Adaptive Evolution Control using Confident Regions for Surrogate-assisted Optimization.** Participants: G. Briffoteaux and N. Melab, external collaborators: M. Mezmaz and D. Tuyttens from Université de Mons (BELGIUM)

  The challenge of the efficient/effective integration of surrogates in the optimization process is to find the best trade-off between the quality (in terms of quality/precision) of the generated solutions and the efficiency (in terms of execution time) of the resolution. In [22], we have investigated the evolution control that alternates between the real function (simulator) and the surrogate within the multi-objective optimization process. We propose an adaptive evolution control mechanism based on the distance-based concept of confident regions (hyperspheres). The approach has been integrated into an ANN-assisted NSGA-2 and experimented using the ZDT4 multi-modal benchmark function. The reported results show that the proposed approach outperforms two other existing ones.

- **A surrogate model for combinatorial optimization.**
  Participants: B. Derbel and A. Liefooghe, external collaborators: H. Aguirre and K. Tanaka, Shinshu University (JAPAN), S. Verel, Univ. Littoral (FRANCE)

  Extensive efforts so far have been devoted to the design of effective surrogate models for expensive black-box continuous optimization problems. There are, however, relatively few investigations on the development of methodologies for combinatorial domains. In [31], we rely on the mathematical foundations of discrete Walsh functions in order to derive a surrogate model for pseudo-boolean optimization functions. Specifically, we model such functions by means of Walsh expansion. By conducting a comprehensive set of experiments on nk-landscapes, we provide empirical evidence on the accuracy of the proposed model. In particular, we show that a Walsh-based surrogate model can outperform the recently-proposed discrete model based on Kriging.

- **Landscape analysis for multi-objective optimization.**
  Participants: B. Derbel and A. Liefooghe, external collaborators: H. Aguirre and K. Tanaka, Shinshu University (JAPAN); M. López-Ibánez, Univ. Manchester (UK); L. Paquete, Univ. Coimbra, Portugal; S. Verel, Univ. Littoral (FRANCE)

  Pareto local optimal solutions (PLOS) are believed to highly influence the dynamics and the performance of multi-objective optimization algorithms, especially those based on local search and Pareto dominance. In [28], we introduce a PLOS network (PLOS-net) model as a step toward the fundamental understanding of multi-objective landscapes and search algorithms. Using a comprehensive set of instances, PLOS-nets are constructed by full enumeration, and selected network features are further extracted and analyzed with respect to instance characteristics. A correlation and regression analysis is then conducted to capture the importance of the PLOS-net features on the runtime and effectiveness of two prototypical Pareto-based heuristics. In particular, we are able to provide empirical evidence for the relevance of the PLOS-net model to explain algorithm performance.

Additionally, we know that local search algorithms naturally stop at a local optimal set (LO-set) under given definitions of neighborhood and preference relation among subsets of solutions, such as set-based dominance relation, hypervolume or epsilon indicator. It is an open question how LO-sets under different set preference relations relate to each other. In [29], we report an in-depth experimental analysis on multi-objective nk-landscapes. Our results reveal that, whatever the preference relation, the number of LO-sets typically increases with the problem non-linearity, and decreases with the number of objectives. We observe that strict LO-sets of bounded cardinality under set-dominance are LO-sets under both epsilon and hypervolume, and that LO-sets under hypervolume are LO-sets under set-dominance, whereas LO-sets under epsilon are not. Nonetheless, LO-sets under set-dominance are more similar to LO-sets under epsilon than under hypervolume. These findings have important implications for multi-objective local search. For instance, a dominance-based approach with bounded archive gets more easily trapped and might experience difficulty to identify an LO-set under epsilon or hypervolume. On the contrary, a hypervolume-based approach is expected to perform more steps before converging to better approximations.

7.3. Large scale GPU-centric optimization

Participants: J. Gmys, T. C. Pessoa and N. Melab, external collaborators: M. Mezmaz, D. Tuyttens from University of Mons (BELGIUM) and F.H. De Carvalho Junior from Universidade Federal Do Cearà (BRAZIL)

Nowadays, accelerator-centric architectures offer orders-of-magnitude performance and energy improvements. The interest of those parallel resources has been recently accentuated by the advent of deep learning making them definitely key-building blocks of modern supercomputers. During the year 2018, in collaboration with A. Zomaya (The Univ. of Sydney) and I. Chakroun (IMEC, Leuven) N. Melab has (guest-)edited a special issue on this hot topic (editorial in [16]). In addition, we have put the focus on the investigation of these specific devices within the context of parallel optimization. In the following, two major contributions are reported: (1) Many-core Branch-and-Bound for GPU accelerators and MIC coprocessors; (2) Cuda Dynamic Parallelism (CDP) for backtracking.

- **Many-core Branch-and-Bound for GPU accelerators and MIC coprocessors.** Solving large optimization problems results in the generation of a very large pool of subproblems and the time-intensive evaluation of their associated lower bounds. Generating and evaluating those subproblems on coprocessors raises several issues including processor-coprocessor data transfer optimization, vectorization, thread divergence, etc. In [15], [32], we have investigated the offload-based parallel design and implementation of B&B algorithms for coprocessors addressing these issues. Two major many-core architectures are considered and compared: Nvidia GPU and Intel MIC. The proposed approaches have been experimented using the Flow-Shop scheduling problem and two hardware configurations equivalent in terms of energy consumption: Nvidia Tesla K40 and Intel Xeon Phi 5110P. The reported results show that the GPU-accelerated approach outperforms the MIC offload-based one even in its vectorized version. Moreover, vectorization improves the efficiency of the MIC offload-based approach with a factor of two.
- **Dynamic Configuration of CUDA Runtime Variables for CDP-based Divide-and-Conquer Algorithms.** CUDA Dynamic Parallelism (CDP) is an extension of the GPGPU programming model proposed to better address irregular applications and recursive patterns of computation. However, processing memory-demanding problems by using CDP is not straightforward, because of its particular memory organization. We have proposed in [23] (extension of [13]) an algorithm to deal with such an issue which dynamically calculates and configures the CDP runtime variables and the GPU heap on the basis of an analysis of the partial backtracking tree. We have implemented the algorithm for solving permutation problems and experimented on two test-cases: N-Queens and the Asymmetric Travelling Salesman Problem. The proposed algorithm allows different CDP-based backtracking from the literature to solve memory-demanding problems, adaptively with respect to the number of recursive kernel generations and the presence of dynamic allocations on GPU.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Our current industrial contracts and granted projects are completely at the heart of the BONUS project. They are summarized in the following.

- *Beckman & Coulter (2015-2018, California, USA)*: the goal of this contract is the strategic and operational planning of large medical laboratories (Phd of S. Faramarzi-Oghani). More exactly, the focus is put on the multi-objective modeling and solving of large (e.g. dozens of thousands of medical test tubes to be analyzed) strategic, tactical and operational problems such as the layout design, machine selection and confguration, assignment and scheduling. The project deals also with the coupling between optimization and simulation for performance assessment.
- *EDF (2015-2019, Paris)*: this project deals with demand-side management in smart grids with EDF, a major electrical power player in France. The Energy Management System (EMS) in the home receives the market and system signals and controls the loads, Heating, Ventilation and Air Conditioning systems (HVAC), storages and local generation units according to the user preferences. A large number of home users and appliances and several conflicting objectives have to be considered.
- *ONERA & CNES (2016-2020, Paris)*: the focus of this project with major European players in vehicle aerospace is put on the design of aerospace vehicles, a high-dimensional expensive multidisciplinary problem. Such problem needs the use of the research lines of BONUS to be tackled effectively and efficiently. Two jointly supervised Phd students (J. Pelamatti and A. Hebbal) are involved in this project.
- *In contact with EXOTEC (2018-2019, Lille)*: This project deals with the optimization of logistics flows of robots. More exactly, the problem consists in efficient complex scheduling without collision of thousands of missions realized by a fleet of dozens of robots and several operators in a 3D logistics warehouse.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- *CPER Data (2015-2019)*: in this project, that promotes research and software development related to advanced data science, the BONUS team is the scientific leader (N. Melab) of one of the three research lines of the project “Optimization and High-Performance Computing”. In this context, a two-year (2018-2019) engineer (J-Y. Ji) is supported to develop a software demonstrator on decomposition-based big optimization. In addition, the team is co-leader of the workpackage/lever “Research infrastructures” related to the Grid’5000 nation-wide experimental testbed. This allowed to extend the testbed at Lille with a GPU-powered cluster highly important for the BONUS project. In addition, two engineers have been hired for the system & network administration of the testbed, user support and development.
- *CPER ELSAT (2015-2019)*: in this project, focused on ecomobility, security and adaptability in transport, the BONUS team is involved in the transversal research line: planning and scheduling of maintenance logistics in transportation. The team got support for a one-year (2017-2018) post-doc position (M. Rahimi).

9.2. National Initiatives

9.2.1. ANR

- *Bilateral ANR/RGC France/Hong Kong PRCI* (2016-2021), “Big Multi-objective Optimization” in collaboration with City University of Hong Kong

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 optimization of several, often time-consuming and conflicting objectives. For example, they require the maximization of the product quality while minimizing the production cost, and rely on demanding numerical simulations in order to assess the objectives. These, so-called multi-objective optimization problems can be solved more efficiently if parallelization is used to execute the simulations simultaneously and if the simulations are partly replaced by accurate surrogate models.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

Program: COST CA15140

Project acronym: ImAppNIO

Project title: Improving applicability of nature-inspired optimization by joining theory and practice

Duration: 2016-2019

Coordinator: Thomas Jansen

Abstract: The main objective of the COST Action is to bridge this gap and improve the applicability of all kinds of nature-inspired optimisation methods. It aims at making theoretical insights more accessible and practical by creating a platform where theoreticians and practitioners can meet and exchange insights, ideas and needs; by developing robust guidelines and practical support for application development based on theoretical insights; by developing theoretical frameworks driven by actual needs arising from practical applications; by training Early Career Investigators in a theory of nature-inspired optimisation methods that clearly aims at practical applications; by broadening participation in the ongoing research of how to develop and apply robust nature-inspired optimisation methods in different application areas.

9.3.3. Collaborations with Major European Organizations

MARO: University of Mons (BELGIUM), Parallel surrogate-assisted optimization, large-scale exact optimization

University of Ceara (BRAZIL), Large-scale GPU-accelerated tree-based optimization

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 Coimbra and University of Lisbon (PORTUGAL), Exact and heuristic multi-objective search

University of Manchester (UNITED KINGDOM), Local optimality in multi-objective optimization

University of Elche and University of Murcia (SPAIN), Matheuristics for DEA

9.4. International Initiatives

9.4.1. Inria International Labs

International Laboratory for Research in Computer Science and Applied Mathematics

Associate Team involved in the International Lab:

9.4.1.1. 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.

9.4.1.2. Other IL projects

Title: Frontiers in Massive Optimization and Computational Intelligence

International Partner (Institution - Laboratory - Researcher): Shinshu University (JAPAN)

Start year: 2017

See also: https://sites.google.com/view/lia-modo/

Abstract: The aim of MODO is to federate French and Japanese researchers interested in the dimensionality, heterogeneity and expensive nature of massive optimization problems. The team receives a yearly support for international exchanges and shared manpower (joint PhD students).

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

Title: Three-fold decomposition in multi-objective optimization (D^3MO)

International Partner (Institution - Laboratory - Researcher): University of Exeter, UK

Start year: 2018

9.4.3. Inria International Partners

9.4.3.1. Informal International Partners

- Collaboration with Université de Mons (UMONS). The collaboration consists mainly in the joint supervision of two Phds (M. Gobert and G. Briffoteaux)
- University of Elche, Spain

9.4.4. Participation in Other International Programs

Title: **Evolutionary many-objective optimization: application to smart cities and engineering design**

International Partner (Institution - Laboratory - Researcher): CINVESTAV-IPN (MEXICO)

Start year: 2016

Abstract: The project is co-funded by ECOS Nord (FRANCE) and ANUIES (MEXICO). Abstract to be extended ...

Title: **Bridging the gap between exact methods and heuristics for multi-objective search (MOCO-Search)**

International Partner (Institution - Laboratory - Researcher): University of Coimbra and University of Lisbon, Portugal

Start year: 2018

Website: http://sites.google.com/view/moco-search/

Abstract: This international project for scientific cooperation (PICS), funded by CNRS and FCT, aims to fill the gap between exact and heuristic methods for multi-objective optimization. The goal is to establish the link between the design principles of exact and heuristic methods, to identify features that make a problem more difficult to be solved by each method, and to improve their performance by hybridizing search strategies. Special emphasis is given to rigorous performance assessment, benchmarking, and general-purpose guidelines for the design of exact and heuristic multi-objective search.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Kiyoshi Tanaka, Shinshu University (JAPAN), March 2018 and November 2018
- Hernan Aguirre, Shinshu University (JAPAN), Invited Professor Univ Lille, from March 2018 until April 2018
- Kalyan Deb, University of Michiga (USA), Oct 2018
- Rachid Ellaia, EMI University of Rabat, Morocco, April 2018

9.5.1.1. Internships

- Alexandre Jesus, University of Coimbra (Portugal)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- E-G. Talbi (Conference program chair): Intl. Conf. on Metaheuristics and Nature Inspired Computing (META'2018), Marrakech, Morocco, Oct. 2018.
- N. Melab (Workshop co-chair): Intl. Workshop on the Synergy of Parallel Computing, Optimization and Simulation (HPCS/PaCOS'2018), Orléans, FRANCE, Jul. 16-20, 2018.
- E-G. Talbi (General Chair): 8th Intl. Conf. on Bioinspired Optimization Methods and their Applications (BIOMA'2018), Paris, FRANCE, May 16-18, 2018.
- E-G. Talbi (Workshop co-chair): Intl. Workshop on Optimization and Learning: Challenges and Applications (OLA'2018), Alicante, SPAIN, Feb. 26-28, 2018.
- E-G. Talbi (steering committee): 8th IEEE Workshop Parallel Distributed Computing and Optimization (IPDPS/PDCO'2018), Vancouver, CANADA, May 21-25, 2018.
- N. Melab: Chair of 4 simulation and HPC-related seminars at Université de Lille, Oct-Dec. 2018 (CENAERO-BELGIUM, IBM, UCL-BELGIUM, ONERA).
- B. Derbel and A. Liefooghe (workshop co-chairs): 1st International Workshop on Computational Intelligence for Massive Optimization (CIMO 2018), Nagano, Japan, July 2018 (with H. Aguirre, K. Tanaka and S. Verel).
- B. Derbel (workshop co-chair): Decomposition Techniques in Evolutionary Optimization (DTEO), workshop at GECCO 2018, Kyoto, Japan, July 2018 (with K. Li, X. Li, S. Zapotecas, Q. Zhang and H. Li).

- A. Liefooghe (workshop co-chair): Landscape-aware heuristic search (LAHS), workshop at GECCO 2018, Kyoto, Japan, July 2018 (with N. Veerapen, S. Verel and G. Ochoa).
- B. Derbel (special session co-chair): Advances in Decomposition-based Evolutionary Multi-objective Optimization (ADEMO), special session at WCCI/CEC 2018, Rio, Brazil, July 2018 (with S. Zapotecas and Q. Zhang).
- B. Derbel and A. Liefooghe (special session co-chairs): Multi-/many-objective optimization and learning, special session at BIOMA 2018, Paris, France, May 2018 (with H. Aguirre, B. Filipič, T. Tušar, and S. Verel).

10.1.1.2. Member of the Organizing Committees

- N. Melab and E-G. Talbi: Synergy Summer School on Efficient Multi-objective Optimization, Ljubljana, Slovenia, Aug. 27-31, 2018.
- E-G. Talbi: The first international Metaheuristics Summer School – MESS 2018, Acireale-Sicily, Italy, Jul. 21-25, 2018.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- N. Melab (Program co-chair): 8^{th} Intl. Conf. on Bioinspired Optimization Methods and their Applications (BIOMA'2018), Paris, France, May 16-18, 2018.
- A. Liefooghe (program co-chair): EvoCOP 2018: 18th European Conference on Evolutionary Computation in Combinatorial Optimisation (Parma, Italy, 2018).
- A. Liefooghe (proceedings chair): GECCO 2018: Genetic and Evolutionary Computation Conference (Kyoto, Japan, 2018).

10.1.2.2. Member of the Conference Program Committees

- IEEE Congress on Evolutionary Computation (CEC), Rio de Janeiro, BRAZIL, Jul. 8-13, 2018.
- The ACM Genetic and Evolutionary Computation Conference (GECCO), Kyoto, JAPAN, July 15-19, 2018.
- The 2018 International Conference on High Performance Computing & Simulation (HPCS), Orleans, FRANCE, July 16–20, 2018.
- IEEE Intl. Workshop on Parallel/Distributed Computing and Optimization (IPDPS/PDCO), Vancouver, Britich Columbia, CANADA, May 21-25, 2018.
- IEEE Intl. on High-Performance Optimization in Industry (HPOI), Ljubljana, SLOVENIA, October 8, 2018.
- Colloque sur l'Optimisation et les Systèmes d'information (COSI), Oran, ALGERIE, Oct. 22-24, 2018.
- The 4^{th} Intl. Conf. on Cloud Computing Technologies and Applications (CloudTech), Brussels, BELGIUM, Nov. 26-28, 2018.
- EvoCOP'2018, 18^{th} European Conference on Evolutionary Computation in Combinatorial Optimization, Parma, ITALY, April 4–6, 2018.
- 15th International Conference on Parallel Problem Solving from Nature (PPSN), Coimbra, Portugal, September 2018
- Genetic and Evolutionary Computation Conference (GECCO), Kyoto, Japan, July 2018.
- IEEE Congress on Evolutionary Computation (WCCI-CEC), Rio, Brazil, July 2018.
- 18th European Conference on Evolutionary Computation in Combinatorial Optimisation (EvoCOP), Parma, Italy, April 2018.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- N. Melab: Guest and Managing 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), Vol. 112, 2018.
- P. Korosec, N. Melab and E-G. Talbi. Guest Editor of LNCS Proceedings of 8^{th} Intl. Conf., BIOMA 2018, Paris, France, May 16-18, 2018. *Springer Lecture Notes in Computer Science (LNCS)*, Vol. 10835, 2018.
- B. Derbel: Associate Editor, IEEE Transactions on Systems, Man and Cybernetics: Systems (IEEE).
- A. Liefooghe, M. López-Ibánez: Editors of LNCS Proceedings of the 18th European conference on evolutionary computation in combinatorial optimization (EvoCOP 2018), Lecture Notes in Computer Science (LNCS), vol. 10782, Parma, Italy, 2018.
- E-G. Talbi: Co-editor (with C. Ribeiro) of a special issue in International Transactions on Operational Research (ITOR) on Optimization and Learning, 2018.

10.1.3.2. Reviewer - Reviewing Activities

- Journal of Heuristics (Springer)
- IEEE Transactions on Parallel and Distributed Systems
- IEEE Transactions on Cybernetics
- Evolutionary Computation (MIT)

10.1.4. Invited Talks

- N. Melab: High-performance Computing, Invited Speaker (1h30), Synergy Summer School, Ljubljana, SLOVENIA, Aug. 28^{th}, 2018.
- N. Melab: Introduction to High-performance Computing, Invited Tutorial (1h40), the 7^{th} Intl. Conf. on Metaheuristics and Nature Inspired Computing (META'18), Marrakech, MOROCCO, Oct. 27-31, 2018.
- E-G. Talbi: Bridging the gap between metaheuristics and machine learning, Invited seminar, PUCV Universidad, Santiago, Chile, Mar 2018.
- E-G. Talbi: How machine learning can help metaheuristics, Invited keynote, LOPAL'2018 International Conference on Learning and Optimization Algorithms: Theory and Applications, Rabat, Marrakech, May 2018.
- E-G. Talbi: Synergy between metaheuristics and machine learning, Tutorial, BIOMA'2018 International Conference on Bioinspired Optimization and their Applications, Paris, France, May 2018.
- E-G. Talbi: Parallel and distributed evolutionary algorithms, Invited tutorial, IEEE WCCI World Congress on Computational Intelligence, Rio de Janeiro, Brazil, July 2018.
- E-G. Talbi: Optimization for machine learning, Invited seminar, Universidad Elche, Spain, Dec 2018.

10.1.5. Leadership within the Scientific Community

- N. Melab: scientific leader of Grid'5000 (https://www.grid5000.fr) at Lille, since 2004
- 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
- A. Liefooghe: co-secretary of the association "Artificial Evolution" (EA)

10.1.6. Scientific Expertise

- N. Melab: Reviewer expert for AAPG ANR, CES 23 (B.7, Axe 4), JCJC, FRANCE, 2018

- N. Melab: Member of the advisory committee for the IT and maganement engineer training at Faculté Polytechnique de Mons, BELGIUM

10.1.7. Research Administration

- N. Melab: Member of the steering committee of “Maison de la Simulation” at Université de Lille
- E-G. Talbi, Coordinator of the International Relationships of Inria Lille Nord Europe

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

- International Master lecture: N. Melab, Supercomputing, 24h ETD, M2, Université de Lille, FRANCE
- Master lecture: N. Melab, Operations Research, 72h ETD, M1, Université de Lille, FRANCE
- Master leading: N. Melab, Co-head (with B. Merlet) of the international Master 2 of advanced scientific computing, Université de Lille, FRANCE
- Licence: A. Liefooghe, Algorithmic and Data structure, 36h ETD, L2, Université de Lille, FRANCE
- Licence: A. Liefooghe, Algorithmic for Operations Research, 36h ETD, L3, Université de Lille, FRANCE
- Master: A. Liefooghe, Databases, 30h ETD, M1, Université de Lille, FRANCE
- Master: A. Liefooghe, Advanced Object-oriented Programming, 53h ETD, M2, Université de Lille, FRANCE
- Master: A. Liefooghe, Combinatorial Optimization, 10h ETD, M2, Université de Lille, FRANCE
- Master: A. Liefooghe, Multi-criteria Decision Aid and Optimization, 25h ETD, M2, Université de Lille, FRANCE
- Master leading: A. Liefooghe, superviser of the Master 2 MIAGE IPI-NT
- Master: Bilel Derbel, Combinatorial Optimization, 35h, M2, Université de Lille, FRANCE
- Master: Bilel Derbel, Grid Computing, 16h, M2, Université de Lille, FRANCE
- Master: Bilel Derbel, Parallel and Distributed Programming, 35h, M1, Université de Lille, FRANCE
- Master: Bilel Derbel, Algorithms and Applications, 28h, M1, Université de Lille, FRANCE
- Engineering school: El-Ghazali Talbi, Advanced optimization, 36h, Polytech’Lille, Université de Lille, FRANCE
- Engineering school: El-Ghazali Talbi, Data mining, 36h, Polytech’Lille, Université de Lille, FRANCE
- Engineering school: El-Ghazali Talbi, Operations research, 60h, Polytech’Lille, Université de Lille, FRANCE
- Engineering school: El-Ghazali Talbi, Graphs, 25h, Polytech’Lille, Université de Lille, FRANCE
- Master leading: B. Derbel, head of the Master MIAGE, Université de Lille, FRANCE
- Licence: O. Abdelkafi, Computer Science, 46.5 ETD, L1, Université de Lille, FRANCE
- Licence: O. Abdelkafi, Web Technologies, 36 ETD, L1, Université de Lille, FRANCE
- Licence: O. Abdelkafi, Unix system introduction, 6 ETD, L2, Université de Lille, FRANCE
- Licence: O. Abdelkafi, Web Technologies, 24 ETD, L2 S3H, Université de Lille, FRANCE
- Licence: O. Abdelkafi, object-oriented programming, 36 ETD, L2, Université de Lille, FRANCE
- Licence: O. Abdelkafi, Relational Databases, 36h ETD, L3, Université de Lille, FRANCE
- Licence: O. Abdelkafi, Algorithmic for Operations Research, 36h ETD, L3, Université de Lille, FRANCE

10.2.2. Supervision

- PhD defended: Sohrab Faramarzi, Optimization of medical laboratories, Defended on Dec. 17th, El-Ghazali Talbi
- PhD in progress: Z. Garroussi, Demand side management in smart grids: Multi-objective models, El-Ghazali Talbi and Rachid Ellaia (EMI, Morocco)
- PhD in progress: J. Pelamatti, Multi-disciplinary design of aerospace vehicles, Jan 2017, El-Ghazali Talbi
- PhD in progress: Ali Hebbal, Surrogate-assisted multi-objective evolutionary algorithms, Oct 2017, El-Ghazali Talbi and Nouredine Melab
- PhD in progress (cotutelle): Maxime Gobert, Surrogate-assisted multi-objective evolutionary algorithms, Oct 2017, Nouredine Melab (Université de Lille) and Daniel Tuyttens (Université de Mons, BELGIUM)
- PhD in progress (cotutelle): Guillaume Briffoteaux, Surrogate-assisted multi-objective evolutionary algorithms, Oct 2017, Nouredine Melab (Université de Lille) and Daniel Tuyttens (Université de Mons, BELGIUM)
- PhD in progress: Geoffrey Pruvost, Machine learning and decomposition techniques for large-scale multi-objective optimization, Oct 2018, Bilel Derbel and Arnaud Liefooghe
- PhD in progress: Nicolas Berveglieri, Meta-models and machine learning for massive expensive optimization, Oct 2018, Bilel Derbel and Arnaud Liefooghe
- PhD in progress: Alexandre Jesus, Algorithm selection in multi-objective optimization, Bilel Derbel and Arnaud Liefooghe (University of Lille), Luís Paquete (University of Coimbra, PORTUGAL)

10.2.3. Juries

- N. Melab: PhD thesis of Yahya Al Dhuraibi, Flexible Framework for Elasticity in Cloud Computing, Université de Lille (FRANCE), Dec. 10th 2018.
- N. Melab: PhD thesis of Muhammad Umer Wasim, Design and Implementation of Legal Protection for Trade Secrets in Cloud Brokerage Architectures relying on Blockchains, University of Bologna (ITALY), Apr. 2018.
- N. Melab: PhD thesis of Maruf Ahmed, On Improving The Performance and Resource Utilization of Consolidated Virtual Machines: Measurement, Modeling, Analysis and Prediction, The University of Sydney (AUSTRALIA), Aug. 2018.
- B. Derbel: PhD thesis of Christopher Jankee, Optimisation par métaheuristique adaptative distribuée en environnement de calcul parallèle, Université du Littoral Côte d'Opale (FRANCE), Aug. 2018.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- N. Melab: Nominated again as Chargé de Mission of High Performance Computing and Simulation at Université de Lille, since 2010.
- E-G. Talbi: International relations coordinator for Inria Lille Nord Europe, since 2016.
- N. Melab. Member of the Working Group on software and technological demonstrators, since end 2017.

11. Bibliography

Major publications by the team in recent years

[1] O. ABDELKAFI, L. IDOUMGHAR, J. LEPAGNOT.*A Survey on the Metaheuristics Applied to QAP for the Graphics Processing Units*, in "Parallel Processing Letters", 2016, vol. 26, n^{o} 3, p. 1–20

[2] A. BENDJOUDI, N. MELAB, E. TALBI.*FTH-B&B: A Fault-Tolerant HierarchicalBranch and Bound for Large ScaleUnreliable Environments*, in "IEEE Trans. Computers", 2014, vol. 63, n^{o} 9, p. 2302–2315

[3] S. CAHON, N. MELAB, E. TALBI.*ParadisEO: A Framework for the Reusable Design of Parallel and Distributed Metaheuristics*, in "J. Heuristics", 2004, vol. 10, n^{o} 3, p. 357–380

[4] F. DAOLIO, A. LIEFOOGHE, S. VÉREL, H. E. AGUIRRE, K. TANAKA.*Problem Features versus Algorithm Performance on Rugged Multiobjective Combinatorial Fitness Landscapes*, in "Evolutionary Computation", 2017, vol. 25, n^{o} 4

[5] B. DERBEL.*Contributions to single- and multi- objective optimization: towards distributed and autonomous massive optimization*, in "HDR dissertation, Université de Lille", 2017

[6] B. DERBEL, A. LIEFOOGHE, Q. ZHANG, H. E. AGUIRRE, K. TANAKA.*Multi-objective Local Search Based on Decomposition*, in "Parallel Problem Solving from Nature - PPSN XIV - 14th International Conference, Edinburgh, UK, September 17-21, 2016, Proceedings", 2016, p. 431–441

[7] 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", 2017, vol. 29, n^{o} 9

[8] A. LIEFOOGHE, B. DERBEL, S. VÉREL, H. E. AGUIRRE, K. TANAKA.*Towards Landscape-Aware Automatic Algorithm Configuration: Preliminary Experiments on Neutral and Rugged Landscapes*, in "Evolutionary Computation in Combinatorial Optimization - 17th European Conference, EvoCOP 2017, Amsterdam, The Netherlands, April 19-21, 2017, Proceedings", 2017, p. 215–232

[9] T. V. LUONG, N. MELAB, E. TALBI.*GPU Computing for Parallel Local Search Metaheuristic Algorithms*, in "IEEE Trans. Computers", 2013, vol. 62, n^{o} 1, p. 173–185

[10] A. NAKIB, S. OUCHRAA, N. SHVAI, L. SOUQUET, E. TALBI.*Deterministic metaheuristic based on fractal decomposition for large-scale optimization*, in "Appl. Soft Comput.", 2017, vol. 61, p. 468–485

Publications of the year

Articles in International Peer-Reviewed Journal

[11] L. ASLI, M. AÏDER, E.-G. TALBI.*Solving a Dynamic combinatorial auctions problem by a hybrid metaheuristic based on a fuzzy dominance relation*, in "RAIRO - Operations Research", 2018 [*DOI :* 10.1051/RO/2018051], https://hal.inria.fr/hal-01942418

[12] O. BAHRI, E.-G. TALBI, N. BEN AMOR.*A generic fuzzy approach for multi-objective optimization under uncertainty*, in "Swarm and Evolutionary Computation", June 2018, vol. 40, p. 166-183 [*DOI :* 10.1016/J.SWEVO.2018.02.002], https://hal.inria.fr/hal-01942402

[13] T. CARNEIRO PESSOA, J. GMYS, F. HERON DE CARVALHO JUNIOR, N. MELAB, D. TUYTTENS.*GPU-Accelerated Backtracking Using CUDA Dynamic Parallelism*, in "Concurrency and Computation: Practice and Experience", May 2018, vol. 30, n^{o} 9 [*DOI :* 10.1002/CPE.4374], https://hal.inria.fr/hal-01919514

[14] N. DUPIN, E.-G. TALBI.*Parallel matheuristics for the discrete unit commitment problem with min-stop ramping constraints*, in "International Transactions in Operational Research", 2018, https://hal.inria.fr/hal-01942412

[15] 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", May 2018, vol. 82, 20 [*DOI :* 10.1016/J.FUTURE.2016.12.039], https://hal.inria.fr/hal-01419079

[16] N. MELAB, A. ZOMAYA, I. CHAKROUN.*Parallel optimization using/for multi and many-core high performance computing*, in "Journal of Parallel and Distributed Computing", February 2018, vol. 112, p. 109 - 110 [*DOI :* 10.1016/J.JPDC.2017.11.011], https://hal.inria.fr/hal-01924680

[17] A. NAKIB, L. SOUQUET, E.-G. TALBI.*Parallel fractal decomposition based algorithm for big continuous optimization problems*, in "Journal of Parallel and Distributed Computing", 2018 [*DOI :* 10.1016/J.JPDC.2018.06.002], https://hal.archives-ouvertes.fr/hal-01844420

[18] O. SCHUTZE, C. HERNANDEZ, E.-G. TALBI, J.-Q. SUN, Y. NARANJANI, F.-R. XIONG.*Archivers for the representation of the set of approximate solutions for MOPs*, in "Journal of Heuristics", 2018, https://hal.inria.fr/hal-01942424

[19] E.-G. TALBI.*A unified view of parallel multi-objective evolutionary algorithms*, in "Journal of Parallel and Distributed Computing", May 2018, p. 1-10 [*DOI :* 10.1016/J.JPDC.2018.04.012], https://hal.inria.fr/hal-01942407

International Conferences with Proceedings

[20] O. ABDELKAFI, L. IDOUMGHAR, J. LEPAGNOT, J.-L. PAILLAUD.*The determination of new stable zeolite frameworks using a parallel hybrid genetic algorithm*, in "OLA 2018 - International Workshop on Optimization and Learning: Challenges and Applications", Alicante, Spain, February 2018, p. 1-2, https://hal.archives-ouvertes.fr/hal-01726493

[21] O. BAHRI, E.-G. TALBI.*Dealing with Epistemic Uncertainty in Multi-objective Optimization: A Survey*, in "IPMU 2018 - 17th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems", Cadiz, Spain, June 2018, p. 260-271, https://hal.inria.fr/hal-01942474

[22] G. BRIFFOTEAUX, N. MELAB, M. MEZMAZ, D. TUYTTENS.*An adaptive evolution control based on confident regions for surrogate-assisted optimization*, in "HPCS 2018 - International Conference on High Performance Computing & Simulation", Orléans, France, July 2018, https://hal.archives-ouvertes.fr/hal-01922708

[23] T. CARNEIRO PESSOA, J. GMYS, N. MELAB, F. HERON DE CARVALHO JUNIOR, P. P. P. REBOUÇAS FILHO, D. TUYTTENS.*Dynamic Configuration of CUDA Runtime Variables for CDP-based Divide-and-Conquer Algorithms*, in "VECPAR 2018 - 13th International Meeting on High Performance Computing for Computational Science", São Pedro, Brazil, September 2018, https://hal.inria.fr/hal-01919532

[24] B. DERBEL, A. LIEFOOGHE, Q. ZHANG, S. VEREL, H. AGUIRRE, K. TANAKA.*A set-oriented MOEA/D*, in "GECCO 2018 - Genetic and Evolutionary Computation Conference", Kyoto, Japan, ACM Press, July 2018, p. 617-624 [*DOI :* 10.1145/3205455.3205575], https://hal.archives-ouvertes.fr/hal-01823671

[25] A. HEBBAL, L. BREVAULT, M. BALESDENT, E.-G. TALBI, N. MELAB.*Efficient Global Optimization using Deep Gaussian Processes*, in "CEC 2018 - Congress on Evolutionary Computation", Rio de Janeiro, Brazil, July 2018, https://hal.inria.fr/hal-01919795

[26] A. HEBBAL, L. BREVAULT, M. BALESDENT, E.-G. TALBI, N. MELAB.*Multi-Disciplinary Design Multi-Objective Optimization of Aerospace Vehicles using Surrogate Models*, in "OLA 2018 - International Workshop on Optimization and Learning: Challenges and Applications", Alicante, Spain, February 2018, https://hal.inria.fr/hal-01942467

[27] A. LIEFOOGHE, B. DERBEL, S. VEREL, H. AGUIRRE, K. TANAKA.*Towards Landscape-Aware Automatic Algorithm Configuration: Preliminary Experiments on Neutral and Rugged Landscapes*, in "European Conference on Evolutionary Computation in Combinatorial Optimization", Parma, Italy, Lecture Notes in Computer Science, April 2018, vol. 10782, p. 215-232, https://hal.archives-ouvertes.fr/hal-01824982

[28] A. LIEFOOGHE, B. DERBEL, S. VEREL, M. LÓPEZ-IBÁÑEZ, H. AGUIRRE, K. TANAKA.*On Pareto local optimal solutions networks*, in "International Conference on Parallel Problem Solving from Nature (PPSN 2018)", Coimbra, Portugal, Lecture Notes in Computer Science, Springer, September 2018, vol. 11102, p. 232-244 [*DOI :* 10.1007/978-3-319-99259-4_19], https://hal.archives-ouvertes.fr/hal-01823721

[29] A. LIEFOOGHE, M. LÓPEZ-IBÁÑEZ, L. PAQUETE, S. VEREL.*Dominance, epsilon, and hypervolume local optimal sets in multi-objective optimization, and how to tell the difference*, in "GECCO 2018 - Genetic and Evolutionary Computation Conference", Kyoto, Japan, ACM Press, July 2018, vol. 18, p. 324-331 [*DOI :* 10.1145/3205455.3205572], https://hal.archives-ouvertes.fr/hal-01823666

[30] J. SHI, Q. ZHANG, B. DERBEL, A. LIEFOOGHE, J. SUN.*Parallel Pareto local search revisited – First experimental results on bi-objective UBQP*, in "GECCO 2018 - Genetic and Evolutionary Computation Conference", Kyoto, Japan, ACM Press, July 2018, p. 753-760 [*DOI :* 10.1145/3205455.3205577], https://hal.archives-ouvertes.fr/hal-01920339

[31] S. VEREL, B. DERBEL, A. LIEFOOGHE, H. AGUIRRE, K. TANAKA.*A surrogate model based on Walsh decomposition for pseudo-boolean functions*, in "International Conference on Parallel Problem Solving from Nature (PPSN 2018)", Coimbra, Portugal, Lecture Notes in Computer Science, September 2018, vol. 11102, p. 181-193 [*DOI :* 10.1007/978-3-319-99259-4_15], https://hal.archives-ouvertes.fr/hal-01823725

Scientific Books (or Scientific Book chapters)

[32] N. MELAB, J. GMYS, M. MEZMAZ, D. TUYTTENS.*Many-core Branch-and-Bound for GPU accelerators and MIC coprocessors*, in "High-performance simulation based optimization", T. BARTZ-BEIELSTEIN, B. FILIPIC, P. KOROSEC, E.-G. TALBI (editors), Springer, 2018, https://hal.inria.fr/hal-01924766

[33] J. PELAMATTI, L. BREVAULT, M. BALESDENT, E.-G. TALBI, Y. GUERIN.*How to deal with mixed-variable optimization problems: An overview of algorithms and formulations*, in "Advances in Structural and Multidisciplinary Optimization", Springer, 2018, p. 64-82 [*DOI :* 10.1007/978-3-319-67988-4_5], https://hal.inria.fr/hal-01690446

Books or Proceedings Editing

[34] P. KOROŠEC, N. MELAB, E.-G. TALBI (editors). *Bioinspired Optimization Methods and Their Applications*, Lecture Notes in Computer Science, Springer, Paris, France, 2018, n° 10835, p. XIII - 333, https://hal.inria.fr/hal-01942377

[35] A. LIEFOOGHE, M. LÓPEZ-IBÁÑEZ (editors). *Proceedings of the 18th European conference on evolutionary computation in combinatorial optimization (EvoCOP 2018)*, Lecture Notes in Computer Science, Springer, Parma, Italy, 2018, vol. 10782 [*DOI :* 10.1007/978-3-319-77449-7], https://hal.archives-ouvertes.fr/hal-01920369

Other Publications

[36] Y. MARCA, H. AGUIRRE, S. ZAPOTECAS-MARTÍNEZ, A. LIEFOOGHE, S. VEREL, B. DERBEL, K. TANAKA.*Pareto dominance-based MOEAs on Problems with Difficult Pareto Set Topologies*, July 2018, GECCO 2018 - Genetic and Evolutionary Computation Conference Companion, Poster, https://hal.archives-ouvertes.fr/hal-01823715

[37] H. MONZÓN, H. AGUIRRE, S. VEREL, A. LIEFOOGHE, B. DERBEL, K. TANAKA.*Studying MOEAs Dynamics and their Performance using a Three Compartmental Model*, July 2018, GECCO 2018 - Genetic and Evolutionary Computation Conference Companion, Poster, https://hal.archives-ouvertes.fr/hal-01823709

[38] J. PELAMATTI, L. BRÉVAULT, M. BALESDENT, E.-G. TALBI, Y. GUERIN.*Efficient global optimization of constrained mixed variable problems*, December 2018, https://arxiv.org/abs/1806.03975 - working paper or preprint, https://hal.inria.fr/hal-01942439

References in notes

[39] M. BALESDENT, L. BREVAULT, N. PRICE, S. DEFOORT, R. LE RICHE, N.-H. KIM, R. HAFTKA, N. BÉREND.*Advanced Space Vehicle Design Taking into Account Multidisciplinary Couplings and Mixed Epistemic/Aleatory Uncertainties*, in "Space Engineering: Modeling and Optimization with Case Studies", Modeling and Optimization with Case Studies, Springer Optimization and Its Applications, 2016, vol. 114, p. 1-48 [*DOI :* 10.1007/978-3-319-41508-6_1], https://hal.archives-ouvertes.fr/hal-01475842

[40] B. DERBEL, D. BROCKHOFF, A. LIEFOOGHE, S. VÉREL.*On the Impact of Multiobjective Scalarizing Functions*, in "Parallel Problem Solving from Nature - PPSN XIII - 13th International Conference, Ljubljana, Slovenia, September 13-17, 2014. Proceedings", 2014, p. 548–558

[41] B. DERBEL, A. LIEFOOGHE, G. MARQUET, E. TALBI.*A fine-grained message passing MOEA/D*, in "IEEE Congress on Evolutionary Computation, CEC 2015, Sendai, Japan, May 25-28, 2015", 2015, p. 1837–1844

[42] R. HAFTKA, D. VILLANUEVA, A. CHAUDHURI.*Parallel surrogate-assisted global optimization with expensive functions – a survey*, in "Structural and Multidisciplinary Optimization", 2016, vol. 54(1), p. 3–13

[43] D. JONES, M. SCHONLAU, W. WELCH.*Efficient Global Optimization of Expensive Black-Box Functions*, in "Journal of Global Optimization", 1998, vol. 13(4), p. 455–492

[44] F. SHAHZAD, J. THIES, M. KREUTZER, T. ZEISER, G. HAGER, G. WELLEIN.*CRAFT: A library for easier application-level Checkpoint/Restart and Automatic Fault Tolerance*, in "CoRR", 2017, vol. abs/1708.02030, http://arxiv.org/abs/1708.02030

[45] N. SHAVIT.*Data Structures in the Multicore Age*, in "Communications of the ACM", 2011, vol. 54, n^o 3, p. 76–84

[46] M. SNIR, AL..*Addressing Failures in Exascale Computing*, in "Int. J. High Perform. Comput. Appl.", May 2014, vol. 28, n^{o} 2, p. 129–173

[47] E. TALBI.*Combining metaheuristics with mathematical programming, constraint programming and machine learning*, in "Annals OR", 2016, vol. 240, n^{o} 1, p. 171–215

[48] T. VU, B. DERBEL.*Parallel Branch-and-Bound in multi-core multi-CPU multi-GPU heterogeneous environments*, in "Future Generation Comp. Syst.", 2016, vol. 56, p. 95–109

[49] X. ZHANG, Y. TIAN, R. CHENG, Y. JIN.*A Decision Variable Clustering-Based Evolutionary Algorithm for Large-Scale Many-Objective Optimization*, in "IEEE Trans. Evol. Computation", 2018, vol. 22, n^{o} 1, p. 97–112

Project-Team DEFROST

DEFormable Robotics SofTware

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
Robotics and Smart environments

Table of contents

1. **Team, Visitors, External Collaborators** 53
2. **Overall Objectives** 54
3. **Research Program** 54
   3.1. Introduction 54
   3.2. Objective 1: Accurate model of soft robot deformation computed in finite time 54
   3.3. Objective 2: Model based control of soft robot behavior 55
   3.4. Objective 3: Modeling the interaction with a complex environment 55
   3.5. Objective 4: Soft Robotic Software 55
   3.6. Objective 5: Validation and application demonstrations 55
4. **Application Domains** 56
   4.1. Industry 56
   4.2. Personal and service robotics 56
   4.3. Entertainment industry and arts 56
5. **Highlights of the Year** 58
   5.1.1. Award from the Robotics Society of Japan 58
   5.1.2. Development of a New Open-Source Plugin for SOFA - Model Order Reduction 58
   5.1.3. Echelon III: A compliant manipulator 58
   5.1.4. Collaboration with Allison Okamura’s team at Stanford 60
6. **New Software and Platforms** 60
   6.1. SOFA 60
   6.2. SoftRobots 60
   6.3. Model Order Reduction Plugin for SOFA 61
   6.4. SoftRobots.Inverse 61
7. **New Results** 61
   7.1. Dynamic control of soft robots 61
   7.2. Vision-based force sensing for soft robots 62
   7.3. Fast, generic and reliable control and simulation of soft robots using model order reduction 62
   7.4. FEM-based kinematics and closed-loop control of soft, continuum manipulators 62
   7.5. FEM-based Deformation Control for Dexterous Manipulation of 3D Soft Objects 63
8. **Bilateral Contracts and Grants with Industry** 64
9. **Partnerships and Cooperations** 65
   9.1. Regional Initiatives 65
   9.2. National Initiatives 65
   9.3. European Initiatives 66
   9.4. International Initiatives 66
      9.4.1. Inria International Labs 66
      9.4.2. Inria International Partners 66
         9.4.2.1. Declared Inria International Partners 66
         9.4.2.2. Informal International Partners 66
   9.5. International Research Visitors 66
      9.5.1. Visits of International Scientists 67
      9.5.2. Visits to International Teams 67
         9.5.2.1. Sabbatical programme 67
         9.5.2.2. Research Stays Abroad 67
10. **Dissemination** 67
   10.1. Promoting Scientific Activities 67
      10.1.1. Scientific Events Organisation 67
         10.1.1.1. General Chair, Scientific Chair 67
         10.1.1.2. Member of the Organizing Committees 67

10.1.2. Scientific Events Selection 67
10.1.2.1. Member of the Conference Program Committees 67
10.1.2.2. Reviewer 67
10.1.3. Journal 68
10.1.3.1. Member of the Editorial Boards 68
10.1.3.2. Reviewer - Reviewing Activities 68
10.1.3.3. Tutorials 68
10.1.4. Invited Talks 68
10.1.5. Leadership within the Scientific Community 69
10.1.6. Scientific Expertise 69
10.1.7. Research Administration 69
10.2. Teaching - Supervision - Juries 69
10.2.1. Teaching 69
10.2.2. Supervision 69
10.2.3. Juries 69
10.3. Popularization 70
10.3.1. Interventions 70
10.3.2. Creation of media or tools for science outreach 70
11. Bibliography **70**

Project-Team DEFROST

Creation of the Team: 2015 January 01, updated into Project-Team: 2017 November 01

Keywords:

Computer Science and Digital Science:

A2.3.3. - Real-time systems
A3.1.1. - Modeling, representation
A5.10. - Robotics
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.6. - Optimization
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization

Other Research Topics and Application Domains:

B2.5.1. - Sensorimotor disabilities
B2.7. - Medical devices
B5.1. - Factory of the future
B5.6. - Robotic systems
B5.7. - 3D printing
B9.2. - Art

1. Team, Visitors, External Collaborators

Research Scientists

Christian Duriez [Inria,Team Leader, Senior Researcher, HDR]
Olivier Goury [Inria, Researcher]
Gang Zheng [Inria, Researcher, HDR]

Faculty Members

Jeremie Dequidt [Univ des sciences et technologies de Lille, Associate Professor]
Alexandre Kruszewski [Ecole centrale de Lille, Associate Professor, HDR]

External Collaborators

Piyush Jain [INSERM, from Feb 2018 until Jun 2018]
Fabien Verbrugghe [Ecole centrale de Lille, from May 2018 until Sep 2018]

Technical Staff

Erwan Douaille [Inria, until Feb 2018]
Thor Morales Bieze [Inria]
Meichun Lin [Univ des sciences et technologies de Lille, from Oct 2018]
Damien Marchal [CNRS]
Felix Vanneste [Inria, until Nov 2018]

PhD Students

Walid Amehri [Inria, from Oct 2018]
Eulalie Coevoet [Inria]
Marwa Mohammed Alaa Eldean Eldiwiny [Inria, until Oct 2018]
Maxime Thieffry [Univ Polytechnique des Hauts de France]
Felix Vanneste [Inria, from Dec 2018]
Zhongkai Zhang [Inria]

Post-Doctoral Fellows
Stefan Escaida Navarro [Inria]
Thomas Morzadec [Inria]

Administrative Assistant
Anne Rejl [Inria]

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 parameter 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 using 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 objective 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 fragile 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 validation 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 objectives of the scientific program.

We will build our own experimental soft robots for the validation of objectives 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 . A 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. In this area, using soft materials are particularly convenient as they provide a mass/carried load ratio several orders higher than traditional robots, highly decreasing the kinetic energy and so increasing the motion speed allowed in presence of humans. Moreover, the technology allows more efficient and ergonomic wearable robotic devices, opening the options for exo-skeletons. This remains to be put in place, but it can 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, 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 is particularly not comfortable. These orthosis will be ideally personalized 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 (see Up-Tex) and with the strategic plan of Inria (Improving Rehabilitation and Autonomy).

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 (Disney’s audio-animatronic) and performing arts. We believe that soft robots could be a good support for art. We are pursuing the collaboration with the artist Jonathan Pepe (see https://jonathan-pepe.com/Haruspices).

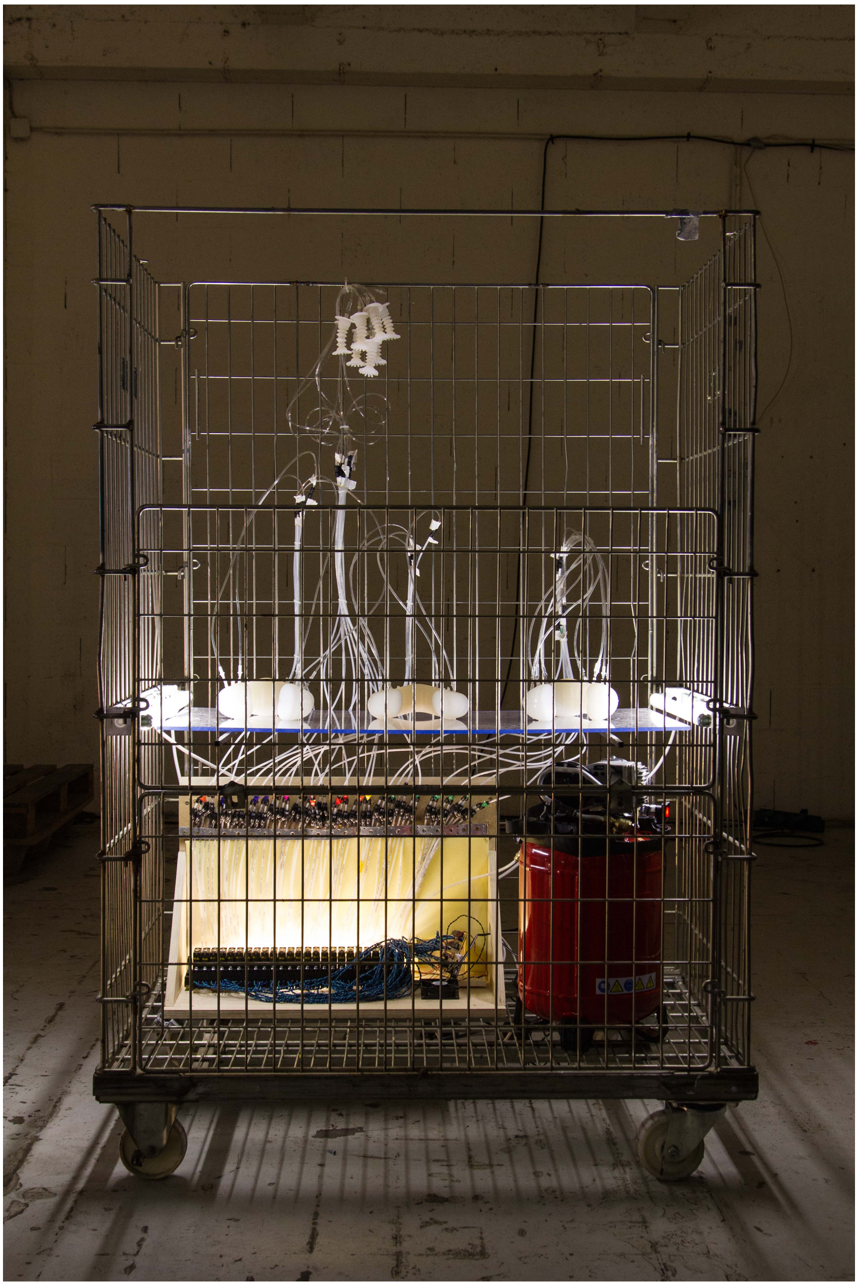

Figure 1. Exobiote project.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Award from the Robotics Society of Japan

We received a best paper award from the *Robotics Society of Japan* for the paper entitled "Software toolkit for modeling, simulation, and control of soft robots" that have been published in the *Advanced Robotics* journal. This paper presents the SoftRobots plugin as a first unified software framework dedicated to modeling, simulation and control of soft robots.

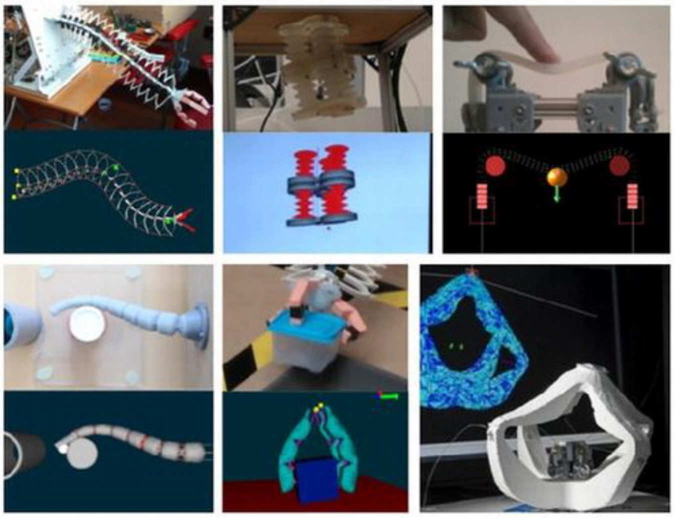

Figure 2. A unified software framework dedicated to modeling, simulation and control of soft robots [2].

5.1.2. Development of a New Open-Source Plugin for SOFA - Model Order Reduction

The plugin Model Order Reduction (MOR) was developed based on the work of the paper [11]. It allows to reduce a SOFA finite element model to gain simulation speed while keeping a good accuracy. It can be used in the SOFA community not only for robotics, but for any application where computational time is an issue, e.g. medical simulations. It is distributed under the GPL license and is available on github: https://github.com/SofaDefrost/ModelOrderReduction.

The plugin is a combination of C++ and Python Code. The user can define the reduction parameters using a python Script or a Graphical User Interface (GUI).

5.1.3. Echelon III: A compliant manipulator

We have participated to the grand challenge of RobotSoft conference that took place in Linorvo, Italy. We have build a robot dedicated to the manipulation competition and we got the 2nd place. A new version of the robot has been developed for the Inria Showroom, installed at Euratechnology in Lille. This version, equipped with a camera, demonstrates the ability of the robot to perform inspection tasks in a limited workspace. We plan to build a new version in 2019 to use it as a research platform, in particular to test planning and control algorithms.

Figure 3. From a computationally intensive simulation to a surrogate version saving accuracy

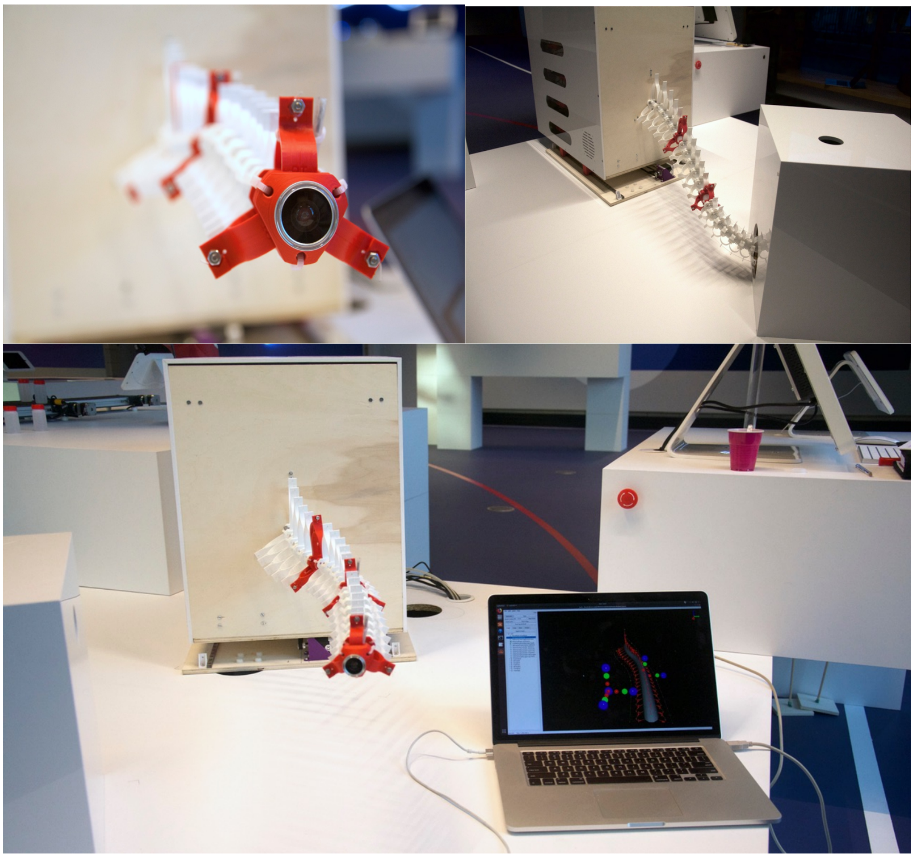

Figure 4. Echelon III in action

5.1.4. Collaboration with Allison Okamura's team at Stanford

This year we had a very close collaboration with the CHARM Lab directed by Allison Okamura at Stanford University. This collaboration resulted in two exchanges: A Stanford PhD student, Margaret Koehler came for 6 months from September 2017 to February 2018 in the team in Lille and Christian Duriez left 7 months of February to August 2018, (thanks to a Fulbright fellowship). We mainly investigated two projects: the haptic rendering on deformable interfaces (A publication in the RAL journal has just been accepted and will be published in 2019) and on the project "Vine Robot" (eversion locomotion). Our teams continue to work on these project. We have also applied to the "Equipe Associée" program.
BEST PAPERS AWARDS :
[2]
E. COEVOET, T. MORALES BIEZE, F. LARGILLIERE, Z. ZHANG, M. THIEFFRY, M. SANZ-LOPEZ, B. CARREZ, D. MARCHAL, O. GOURY, J. DEQUIDT, C. DURIEZ.*Software toolkit for modeling, simulation, and control of soft robots*, in "Advanced Robotics", 2017, https://doi.org/10.1080/01691864.2017.1395362

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.

- Participants: Christian Duriez, François Faure, Hervé Delingette and Stéphane Cotin
- Partner: IGG
- Contact: Stéphane Cotin
- URL: http://www.sofa-framework.org

6.2. SoftRobots

SoftRobots plugin for Sofa
KEYWORDS: Numerical simulations - Problem inverse - Soft robotics
FUNCTIONAL DESCRIPTION: This plugin allows the modeling of deformable robots in the Sofa platform. It allows the modeling of different actuators, such as cable, pneumatic pressure, hydraulics and other simpler types of actuation. It also contains useful tools for animation design or communication with the robot. Coupled with the SoftRobots.Inverse plugin, it also allows the control of these robots. More information can be found on the dedicated website https://project.inria.fr/softrobot/.

- Participants: Christian Duriez, Olivier Goury, Jérémie Dequidt, Damien Marchal, Eulalie Coevoet, Erwan Douaille and Félix Vanneste
- Contact: Christian Duriez
- URL: https://project.inria.fr/softrobot/

6.3. Model Order Reduction Plugin for SOFA

KEYWORDS: Model Order Reduction - Sofa - Finite element modelling
SCIENTIFIC DESCRIPTION: This plugin allows speed-up of SOFA simulations by providing tools to create a reduced version of the SOFA simulation that runs at much higher rates but remains accurate. Starting with a snapshot of the object deformations on a high-dimensional Finite Element mesh, Proper Orthogonal Decomposition (POD) is used to compute a reduced basis of small dimension representing correctly all the possible deformations of the object. The original system describing the object motion is then greatly reduced. To keep numerical efficiency, a hyper-reduction method is used to speed-up the construction of the reduced system.
FUNCTIONAL DESCRIPTION: This plugin allows to dramatically reduce computational time in mechanical simulation in the SOFA framework. A reduced simulation, of much smaller dimension but still accurate is created in an automatic way by the plugin. Building the reduced model may take time, but this operation is made once only. The user can then benefit from a reduced and interactive version of his/her simulation without significant loss of accuracy.

RELEASE FUNCTIONAL DESCRIPTION: This is the first version of the plugin.
NEWS OF THE YEAR: Publication using this plugin accepted dans IEEE Transactions on Robotics

- Participants: Olivier Goury, Félix Vanneste, Christian Duriez and Eulalie Coevoet
- Contact: Olivier Goury
- Publication: Fast, generic and reliable control and simulation of soft robots using model order reduction
- URL: https://project.inria.fr/modelorderreduction/

6.4. SoftRobots.Inverse

KEYWORDS: Sofa - SoftRobots
FUNCTIONAL DESCRIPTION: This plugin builds on the plugin SoftRobots (https://project.inria.fr/softrobot/). Inside the plugin, there is some constraint components that are used to describe the robot (effectors, actuators, sensors). An optimisation algorithm is provided to find the efforts to put on actuators in order to place the robot in a the closest possible configuration than the one described by "effectors", or to a state described by "sensors". This method used to control the soft-robots in the task space is patented.

- Partners: CNRS - Université de Lille - Ecole Centrale de Lille
- Contact: Christian Duriez
- URL: https://project.inria.fr/softrobot.inverse

7. New Results

7.1. Dynamic control of soft robots

The objective is to design a closed-loop strategy to control the dynamics of soft robots. We model the soft robot using the Finite Element Method, which leads to work with large-scale systems that are difficult to control. No unified framework exist to control these robots, especially when considering their dynamics. The main contribution of our work is a reduced order model-based control law, that consists in two main features: a reduced state feedback tunes the performance while a Lyapunov function guarantees the stability of the large-scale closed-loop systems. The method is generic and usable for any soft robot, as long as a FEM model is obtained. Simulation and real robots experiments show that we can control and reduce the settling time of the soft robot and make it converge faster without oscillations to a desired position. It can make the robot converge faster and with reduced oscillations to a desired equilibrium state in the robot's work-space. These results have been presented at the European Control Conference [24] and accepted for publication in Robotics and Automation Letters [8].

7.2. Vision-based force sensing for soft robots

This paper proposes a new framework of external force sensing for soft robots based on the fusion of vision-based measurements and Finite Element Model (FEM) techniques. A precise mechanical model of the robot is built using real-time FEM to describe the relationship between the external forces acting on the robot and the displacement of predefined feature points. The position of these feature points on the real robot is measured using a vision system and is compared with the equivalent feature points in the finite element model. Using the compared displacement, the intensities of the external forces are computed by solving an inverse problem. Based on the developed FEM equations, we show that not only the intensities but also the locations of the external forces can be estimated. A strategy is proposed to find the correct locations of external forces among several possible ones. The method is verified and validated using both simulation and experiments on a soft sheet and a parallel soft robot (both of them have non-trivial shapes). The good results obtained from the experimental study demonstrate the capability of our approach.

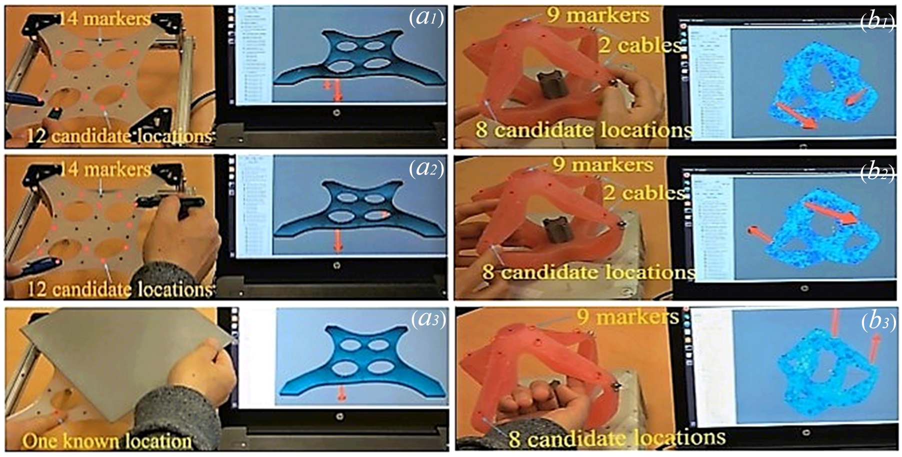

Figure 5. External force sensing for soft objects

7.3. Fast, generic and reliable control and simulation of soft robots using model order reduction

Obtaining an accurate mechanical model of a soft deformable robot compatible with the computation time imposed by robotic applications is often considered as an unattainable goal. This paper should invert this idea. The proposed methodology offers the possibility to dramatically reduce the size and the online computation time of a Finite Element Model (FEM) of a soft robot. After a set of expensive offline simulations based on the whole model, we apply snapshot-proper orthogonal decomposition to sharply reduce the number of state variables of the soft robot model. To keep the computational efficiency, hyper-reduction is used to perform the integration on a reduced domain. The method allows to tune the error during the two main steps of complexity reduction. The method handles external loads (contact, friction, gravity...) with precision as long as they are tested during the offline simulations. The method is validated on two very different examples of FE models of soft robots and on one real soft robot. It enables acceleration factors of more than 100, while saving accuracy, in particular compared to coarsely meshed FE models and provides a generic way to control soft robots.

7.4. FEM-based kinematics and closed-loop control of soft, continuum manipulators

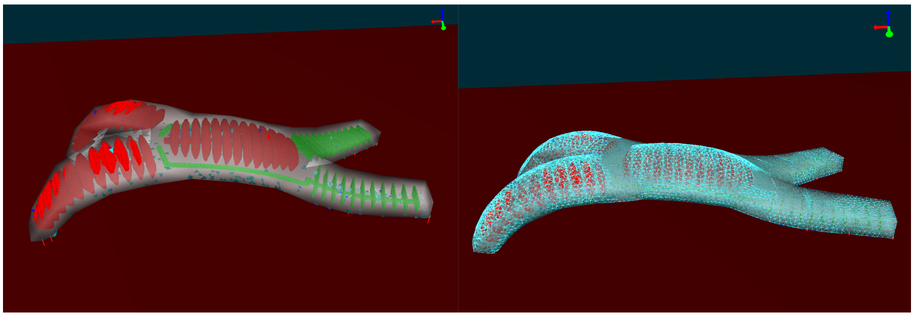

Figure 6. Pneumatic Soft Robot fine simulation versus its surrogate reduced representation manageable in real-time.

This paper presents a modeling methodology and experimental validation for soft manipulators to obtain forward and inverse kinematic models under quasistatic conditions. It offers a way to obtain the kinematic characteristics of this type of soft robots that is suitable for offline path planning and position control. The modeling methodology presented relies on continuum mechanics which does not provide analytic solutions in the general case. Our approach proposes a real-time numerical integration strategy based on Finite Element Method (FEM) with a numerical optimization based on Lagrangian Multipliers to obtain forward and inverse models. To reduce the dimension of the problem, at each step, a projection of the model to the constraint space (gathering actuators, sensors and end-effector) is performed to obtain the smallest number possible of mathematical equations to be solved. This methodology is applied to obtain the kinematics of two different manipulators with complex structural geometry. An experimental comparison is also performed in one of the robots, between two other geometric approaches and the approach that is showcased in this paper. A closed-loop controller based on a state estimator is proposed. The controller is experimentally validated and its robustness is evaluated using Lyapunov stability method.

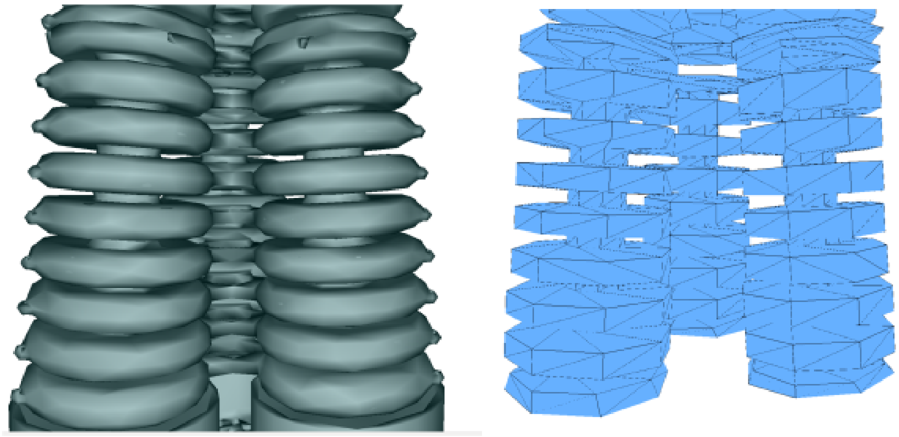

Figure 7. Visual model of the manipulator and the underlying finite element model.

7.5. FEM-based Deformation Control for Dexterous Manipulation of 3D Soft Objects

In this project, that was organize through a collaboration with Fanny Ficuciello from University of Naples and Antoine Petit from Mimesis team in Strasbourg we developed a method for dexterous manipulation of 3D soft objects for real-time deformation control, relying on Finite Element modelling. The goal is to generate proper forces on the fingertips of an anthropomorphic device during in-hand manipulation to produce desired displacements of selected control points on the object. The desired motions of the fingers are computed in real-time as an inverse solution of a Finite Element Method (FEM), the forces applied by the fingertips at the contact points being modelled by Lagrange multipliers. The elasticity parameters of the model are preliminarily estimated using a vision system and a force sensor. Experimental results were shown with an underactuated anthropomorphic hand that performs a manipulation task on a soft cylindrical object.

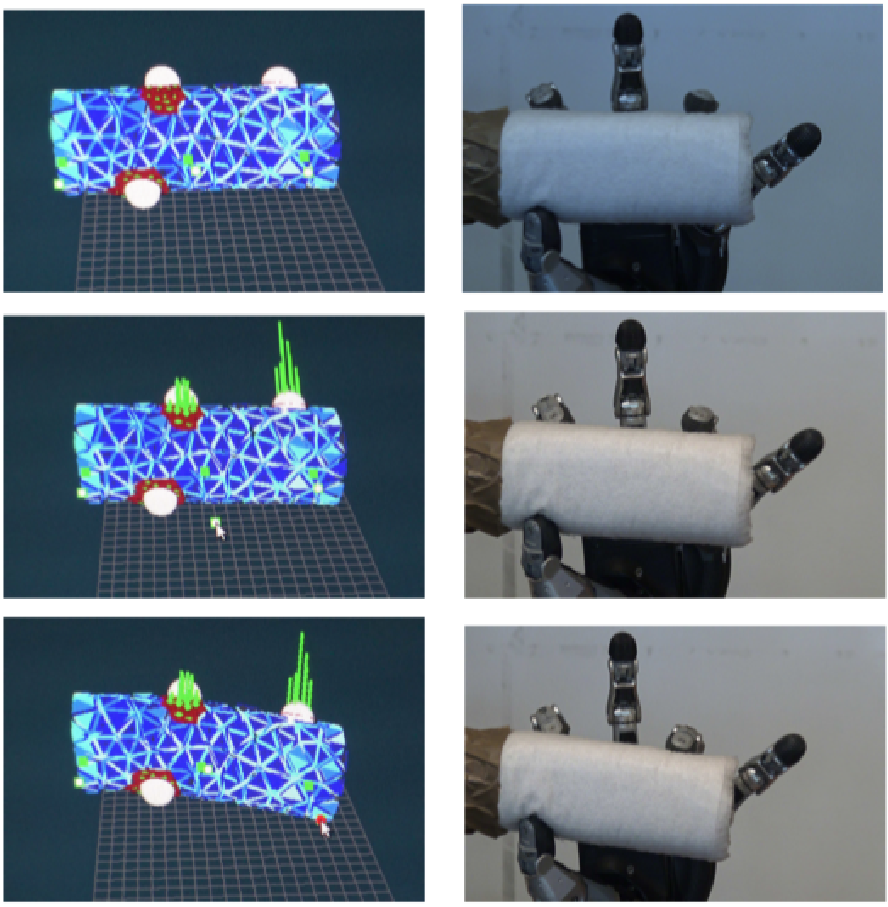

Figure 8. The manipulation of the 3D Soft Object inside the hand is driven by the inverse FEM simulation computed in real-time

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

TDR group is a robotics integrator specialized on optimizing production chains, usually multiplexing robots to perform several activities. Hence, their interest in graspers and the time invested in this activity has been growing within the last years. To improve this aspect, we have been developing together a concept of "universal grasper", based on soft robotics technology and capable of grasping an object with an arbitrary shape, and partially misplaced or misoriented. The prototype developed complies with the specifications and allows for scalability, with flexibility between grasping force and shape tolerance, and the ability for replacing objects without the need of an external vision system. Relying in SOFA for physical simulation, we have validated a prototype, and realize it. An industrial version of the prototype has been realized this year. It will be commercialized next year.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- **INVENTOR** Innovative tool for soft robot design and its application for surgery. This project is financed by **I-Site ULNE EXPAND**, supported by "le programme d'Investissements d'Avenir" and "la Métropole Européenne de Lille". The objective of this project is to develop an innovative tool for the facilitation of soft robot design.
- **COMOROS** Control of deformable robots for surgery Duration april 2017 to march 2020

  Program: FEDER Coordinator: C. Duriez Abstract: Surgical procedures are often carried out using instruments made of stiff materials that interact with delicate biological tissues such as internal organs, blood vessel walls and small cavities. This incompatibility of stiffness is one of the sources of danger in many surgical procedures. The use of robots made of soft materials, also called soft robots, would limit such risks by reducing contact pressures and stress concentrations. Their intrinsic deformability would also increase the ability to manoeuvre in confined spaces. However, the promising concept of using soft robots for surgical procedures cannot be practically implemented, due to the lack of precise modelling and control methods for soft robots. This scientific obstacle, identified as a pending issue by major surveys in this field, becomes particularly challenging when interacting with an environment as complex as the human anatomy. Drawing on our background in soft tissue simulation, contact models, surgical applications and soft robotics, our ambition in this project is to:

  – Develop accurate and generic numerical methods for continuum mechanics, adapted to strong real-time constraints in order to demonstrate the ability to model soft mechatronics systems.
  – Reconsider parametrization methodologies of digital models of the patient anatomy through the observation of mechanical interactions with soft robots via embedded sensors and medical imaging
  – Rethink motion generation and teleoperation control with force feedback so as to be compatible with the large number of degrees of freedom of soft robots and be based on accurate, rapidly-computed deformable models and interaction models.

  The project also targets the development of software with the required performance and features, as well as the experimental validation of models and methods using prototypes in realistic environments.

9.2. National Initiatives

- **SIMILAR** Soft robotIc framework for modeling, simulation and control. This project is supported by **Inria ADT**, and the objective is to design new 3D interactive software to design soft-robots. This new software will be on the top of our existing software stack relying on SOFA for all numerical simulation aspects and 3D rendering aspects.
- **Tremplin ERC** Christian Duriez received a **ANR** grant “tremplin ERC” (150k€) given the result obtained last year on the ERC proposal (evaluated at “grade A”). The project has allowed to allocate new resources on the developments that were presented in this ERC.

9.3. European Initiatives

9.3.1. Collaborations in European Programs, Except FP7 & H2020

Meichun Lin was doing a project belonged to Interreg - 2 Seas Mers Zeeën on Cooperate Brachytherapy(CoBra), it is a 4 years project which gathers the experts from the countries between English Channel and southern North Sea aiming on finding an advance method for curing prostate cancer. The project is divided by several fields which are - MR compatible robot design, radiation dose measurement, needle design and virtual real-time training tool development. Meichun was working on developing virtual real-time training tool with Defrost team. By using SOFA framework to simulate the soft tissue’s deformation and the interaction with needle under the real-time, also with the Image Modelling of MRI, Organs and tissue Modelling and so on and so forth, the 3D rendering became more like the real procedure of the brachytherapy and better for training purpose.

9.4. International Initiatives

9.4.1. Inria International Labs

Christian Duriez realized a geographical mobility as part of the program Inria @ SiliconValley. More details in the Highlights section.

9.4.2. Inria International Partners

9.4.2.1. Declared Inria International Partners

Collaboration with the group of Allison Okamura at Stanford University

Christian Duriez was awarded of a Fulbright Grant for going 7 months (February to August 2018) at Stanford University to work with the group of Allison Okamura. One of a phD student of Stanford, Margaret Koehler, has been awarded of a Chateaubriand Grant for coming 6 months (September 2017 to February 2018) in our Group in Lille. The collaboration was about 2 projects. The first project is haptic rendering with deformable robotics device. The second project is about the modeling and simulation of the “vine robot” that is currently being designed at Stanford.

9.4.2.2. Informal International Partners

- Collaboration with Massachusetts Institute of Technology:

  Maxime Thieffry spent a month in the Distributed Robotics Laboratory, CSAIL, MIT, for a collaboration with Robert Katzschmann and Daniela Rus. This work led to a submission to the soft robotics conference, RoboSoft 2019.
- Collaboration with Dipartimento di Ingegneria Elettrica e delle Tecnologie dell’Informazione, Napoli:

  The project was on the control of manipulation tasks. Using the SoftRobots.Inverse plugin we allowed the control of the shape of a deformable object manipulated by a rigid hand. In the paper [20] we demonstrate the feasibility of the method.

9.5. International Research Visitors

- Prof. Shunjie LI from Nanjing University of Information Science and Technology (China) visited the team from May 10, 2018 to June 30, 2018.

9.5.1. Visits of International Scientists

9.5.1.1. Internships

Margaret Koehler, phD student at Stanford University, has been awarded of a Chateaubriand Grant for coming 6 months (September 2017 to February 2018) in our Group in Lille.

9.5.2. Visits to International Teams

9.5.2.1. Sabbatical programme

This year, Christian Duriez realized a geographical mobility, since he was invited for 7 months in the team of Allison Okamura (Stanford University). He worked on two projects: the creation of deformable haptic interfaces and the mechanical modeling of the “vine robot” (https://www.vinerobots.org). The trip was funded in part by a Fulbright scholarship.

In addition, a doctoral student of their team, Margaret Koehler, makes a stay of 6 months paid by a Chateaubriand Fellowship.

These exchanges are part of the program Inria @ SiliconValley. See the interview of C. Duriez and the interview of M. Koehler.

9.5.2.2. Research Stays Abroad

- Gang ZHENG has visited Nanjing University of Science and Technology (China) for two weeks in July 2018.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- Gang Zheng is a vice-chair of the IFAC Technical Committee “Social Impact of Automation”, International Federation of Automatic Control, TC9.2
- Gang 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

- Gang Zheng, Associate Editor, SIAM CT19, Chengdu, China (SIAM Conference of Control & Its Applications 2019)

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Gang Zheng is IPC member of SIAM-CT19, ICSRT 2019, ICFCTA 2019.

10.1.2.2. Reviewer

Kruszewski:

- ICRA
- ECC
- ACC

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

- Gang Zheng is the member of Editorial Board, Journal of Control Science and Engineering.
- Christian Duriez is member of Editorial Board of IEEE Transactions on Haptics

10.1.3.2. Reviewer - Reviewing Activities

Kruszewki:

- Systems & Control Letters
- Fuzzy Sets and Systems
- IEEE TFS
- Automatica
- IET Control Theory & Applications
- IEEE TCST
- IEEE Transactions on Industrial Informatics
- Nonlinear Analysis: Hybrid Systems

Olivier Goury:

- IEEE Robotics and Automation Letters

Christian Duriez:

- IEEE Robotics and Automation Letters
- IEEE Transactions on Robotics
- ACM Transactions on Graphics
- Computer Methods in Applied Mechanics and Engineering

Thor Morales Bieze:

- MDPI Actuators
- IEEE Transactions on Robotics
- Soft Robotics
- IEEE Robotics & Automation Magazine

10.1.3.3. Tutorials

- A tutorial on the Soft-Robotics plugin for SOFA was presented at the IEEE International Conference on Soft Robotics (Robosoft 2018) in Livorno.
- Tutorials about SOFA Plugins *BeamAdapter*, *Soft-Robots* and *Model Order Reduction* were presented during the 1st International SOFA week 2018 in Strasbourg.

10.1.4. Invited Talks

Christian Duriez:

- Stanford Robotic Seminar
- Harvard School of Engineering
- University of Madison Department of Mechanical Engineering
- Tufts University, Soft robotics seminar
- MIT CSAIL
- Toyota Research (Los Altos, California)
- Séminaire de Robotique, Université Jules Verne d'Amiens

10.1.5. Leadership within the Scientific Community

Christian Duriez has been invited to participate to the NSF-NIST-DARPA- workshop on “Simulation and Machine Learning in Robotics”. The outcome was to have an impact on the future calls of research projects in the field.

10.1.6. Scientific Expertise

Christian Duriez has been expert for evaluation of ANR Projects and is an evaluator of the European project (FET) Hybridheart. Damien Marchal has been expert for evaluation of HCERES.

10.1.7. Research Administration

A. Kruszewski and C. Duriez are members of the Laboratory council (CRIStAL). C. Duriez is vice-president of the “Commission des Emplois de Recherche” of Lille. O. Goury is an elected member of the “Comité de Centre” of Inria Lille.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Ingénieur: Kruszewski, Start & Go Metallophone, 15h, niveau L3, Centrale Lille, France

Ingénieur: Kruszewski, Start & Go Robotique, 46h, niveau L3, Centrale Lille, France

Ingénieur: Kruszewski, Start & Go Conception et environnement, 6h, niveau L3, Centrale Lille, France

Ingénieur: Kruszewski, Modélisation et commande de système: application à la robotique, 17h, niveau L3, Centrale Lille, France

Ingénieur: Kruszewski, Projet recherche, 12h, niveau M1, Centrale Lille, France

Master 2: Christian Duriez, Simulation physique interactive, 24h, Université de Lille

10.2.2. Supervision

PhD in progress: Thieffry, Modélisation et contrôle de robots déformables à grande vitesse, 01/09/2016, C. Duriez, T.M. Guerra, A. Kruszewski

PhD in progress: Zhang, New methods of visual servoing for soft-robots, 01/10/2015, C. Duriez, J. Dequidt, A. Kruszewski

PhD in progress: Coevoet, Méthodes d’optimisation pour la robotique déformable, 01/05/2017, C. Duriez

PhD in Progress: F. Vanneste, Design and simulation of Soft Robots made of mesostructured materials, 01/12/2018, C Duriez, O. Goury

PhD in Progress: Walid Amehri, Property analysis of soft robots, G. Zheng, A. Kruszewski

10.2.3. Juries

Christian Duriez a participé aux jurys suivant:

PhD: Yinoussa Adagolodjo, Couplage de La robotique et de la simulation médicale pour les procédures automatisées, Université de Strasbourg, septembre 2018, M De Matelin et H. Courtecuisse (président du jury)

PhD: Inderjeet Singh, Curve Based Approach for Shape Reconstruction of Continuum Manipulators, R. Merzouki (examinateur)

PhD in progress: Camille Krewcun, Simulation personnalisée de l’implantation de stent à partir de la géométrie coronaire acquise en imagerie OCT, phD start: 1er octobre 2016 , encadrant(s): Laurent Sarry, Émilie Péry

PhD in progress: LAGNEAU Romain, Data-driven models for dexterous manipulation of robots, Maud MARCHAL, Alexandre KRUPPA

10.3. Popularization

- Damien Marchal, Olivier Goury and Christian Duriez participated in “Fête de la science: Opération Chercheurs itinérants”, which involves giving scientific lectures in middle and high schools.
- Olivier Goury presneted a tutorial for the “Journée de l’Enseignement de l’Informatique et de l’Algorithmique (Journée JEIA)”, which is an event for high school teachers willing to learn more about computer science.
- Olivier Goury et Thomas Morzadec took part in the “Journée RIC (Recherche - Innovation - Créativité)”, happening at IMT Lille Douai. A demonstration was made for students of engineering schools and the university of Lille.

10.3.1. Interventions

- National events: Fête de la Science: participations à l’opération “chercheur itinérant”.

10.3.2. Creation of media or tools for science outreach

Développement du site hands-soft-robotics. http://handsonsoftrobotics.lille.inria.fr/

11. Bibliography

Major publications by the team in recent years

[1] E. Coevoet, A. Escande, C. Duriez. *Optimization-based inverse model of soft robots with contact handling*, in "IEEE Robotics and Automation Letters", 2017, vol. 2, n^{o} 3, p. 1413–1419

[2] *Best Paper*
E. Coevoet, T. Morales Bieze, F. Largilliere, Z. Zhang, M. Thieffry, M. Sanz-Lopez, B. Carrez, D. Marchal, O. Goury, J. Dequidt, C. Duriez. *Software toolkit for modeling, simulation, and control of soft robots*, in "Advanced Robotics", 2017, https://doi.org/10.1080/01691864.2017.1395362.

[3] C. Duriez, E. Coevoet, F. Largilliere, T. Morales 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

[4] 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

[5] O. Goury, C. Duriez. *Fast, generic and reliable control and simulation of soft robots using model order reduction*, in "IEEE Transactions on Robotics", 2018, https://hal.archives-ouvertes.fr/hal-01834483

[6] F. Largilliere, 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

[7] T. Morales Bieze, F. Largilliere, A. Kruszewski, Z. Zhang, R. Merzouki, C. Duriez. *FEM-based kinematics and closed-loop control of soft, continuum manipulators*, in "soft robotics", 2018, https://hal.archives-ouvertes.fr/hal-01745625

[8] M. THIEFFRY, A. KRUSZEWSKI, C. DURIEZ, T.-M. GUERRA.*Control Design for Soft Robots based on Reduced Order Model*, in "IEEE Robotics and Automation Letters", 2018

[9] 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

[10] B.-H. DU, A. POLYAKOV, G. ZHENG, Q. QUAN.*Quadrotor trajectory tracking by using fixed-time differentiator*, in "International Journal of Control", April 2018 [*DOI :* 10.1080/00207179.2018.1462534], https://hal.inria.fr/hal-01947365

[11] O. GOURY, C. DURIEZ.*Fast, generic and reliable control and simulation of soft robots using model order reduction*, in "IEEE Transactions on Robotics", 2018, https://hal.archives-ouvertes.fr/hal-01834483

[12] Z. KADER, G. ZHENG, J.-P. BARBOT.*Finite-time and asymptotic left inversion of nonlinear time-delay systems*, in "Automatica", April 2018, https://hal.archives-ouvertes.fr/hal-01781561

[13] K. LANGUEH, G. ZHENG, T. FLOQUET.*Impulsive fixed-time observer for linear time-delay systems*, in "Journal of The Franklin Institute", May 2018, vol. 355, n^{o} 7, p. 3354 - 3366 [*DOI :* 10.1016/J.JFRANKLIN.2018.02.019], https://hal.archives-ouvertes.fr/hal-01932968

[14] M. MBOUP, C. JOIN, M. FLIESS, Y. WANG, G. ZHENG, D. EFIMOV, W. PERRUQUETTI.*Comments on "Differentiator application in altitude control for an indoor blimp robot"*, in "International Journal of Control", 2019 [*DOI :* 10.1080/00207179.2018.1500040], https://hal.inria.fr/hal-01832874

[15] T. MORALES BIEZE, F. LARGILLIERE, A. KRUSZEWSKI, Z. ZHANG, R. MERZOUKI, C. DURIEZ.*FEM-based kinematics and closed-loop control of soft, continuum manipulators*, in "soft robotics", 2018, https://hal.archives-ouvertes.fr/hal-01745625

[16] M. THIEFFRY, A. KRUSZEWSKI, C. DURIEZ, T.-M. GUERRA.*Control Design for Soft Robots based on Reduced Order Model*, in "IEEE Robotics and Automation Letters", January 2019, https://hal.archives-ouvertes.fr/hal-01901031

[17] Y. WANG, G. ZHENG, D. EFIMOV, W. PERRUQUETTI.*Differentiator application in altitude control for an indoor blimp robot*, in "International Journal of Control", March 2018, vol. 91, n^{o} 9, p. 2121 - 2130 [*DOI :* 10.1080/00207179.2018.1441549], https://hal.inria.fr/hal-01889244

[18] Z. ZHANG, J. DEQUIDT, C. DURIEZ.*Vision-Based Sensing of External Forces Acting on Soft Robots Using Finite Element Method*, in "IEEE Robotics and Automation Letters", February 2018, vol. 3, n^{o} 3, p. 1529 - 1536 [*DOI :* 10.1109/LRA.2018.2800781], https://hal.inria.fr/hal-01720645

[19] G. ZHENG, A. POLYAKOV, A. LEVANT.*Delay estimation via sliding mode for nonlinear time-delay systems*, in "Automatica", 2018, vol. 89, p. 266-273, https://hal.inria.fr/hal-01649546

International Conferences with Proceedings

[20] F. Ficuciello, A. Migliozzi, E. Coevoet, A. Petit, C. Duriez. *FEM-based Deformation Control for Dexterous Manipulation of 3D Soft Objects*, in "IROS 2018 - IEEE/RSJ International Conference on Intelligent Robots and Systems", Madrid, Spain, October 2018, https://hal.inria.fr/hal-01889783

[21] C. Gomez, F. J. Bejarano, G. Zheng, T. Floquet. *A class of observers for the observable part of linear systems with commensurate delays*, in "CNCA 2018 - Congreso Nacional de Control Automático", San Luis Potosi, Mexico, October 2018, https://hal.inria.fr/hal-01947394

[22] S. Nateghi, Y. Shtessel, J.-P. Barbot, G. Zheng, L. Yu. *Cyber-Attack Reconstruction via Sliding Mode Differentiation and Sparse Recovery Algorithm: Electrical Power Networks Application*, in "VSS 2018 - 15th International Workshop on Variable Structure Systems", Graz, Austria, July 2018, https://hal.inria.fr/hal-01947385

[23] G. Zheng, S. Li. *Impulsive Observer for Linear Singular Time-Delay Systems*, in "CCC2018 - Proceedings of the 37th Chinese Control Conference", Wuhan, China, July 2018, https://hal.inria.fr/hal-01947376

Conferences without Proceedings

[24] M. Thieffry, A. Kruszewski, T.-M. Guerra, C. Duriez. *Reduced Order Control of Soft Robots with Guaranteed Stability*, in "European Control Conference ECC18", Limassol, Cyprus, June 2018, https://hal.archives-ouvertes.fr/hal-01747433

Other Publications

[25] M. A. ElDiwiny, E. Cattan, C. Duriez. *How To Design Your Polymer Artificial Muscle Actuator/ Sensor?*, May 2018, p. 1-2, ICRA 2018 - IEEE International Conference on Robotics and Automation, Poster, https://hal.archives-ouvertes.fr/hal-01798709

Project-Team FUN

self-organizing Future Ubiquitous Network

RESEARCH CENTER
Lille - Nord Europe

THEME
Networks and Telecommunications

Table of contents

1. Team, Visitors, External Collaborators **77**
2. Overall Objectives **78**
3. Research Program **79**
3.1. Introduction 79
3.2. Vertical Perspective 81
3.3. Horizontal perspective 83
4. Application Domains **83**
5. Highlights of the Year **83**
6. New Software and Platforms **84**
6.1. AspireRFID ALE 84
6.2. ETINODE-CONTIKI-PORT 84
6.3. ETINODE-DRIVERS 84
6.4. EVe-TCF 84
6.5. GOLIATH 85
6.6. IoT-LAB robots 85
6.7. T-SCAN 85
6.8. FIT IoT-Lab 85
7. New Results **86**
7.1. Performance Evaluation, Security, Safety and Verification 86
7.1.1. Performance Evaluation and validation methodology 86
7.1.2. Correlated failures 86
7.1.3. Contiki verification 86
7.2. Alternative communication paradigms 87
7.3. Self-Organization 88
7.3.1. Stable parent selection 88
7.3.2. Bayesian communications 88
7.3.3. Multi-technology self-organization 88
7.3.4. Heterogeneous Self-organizing (smart) Things 89
7.4. Smart Grids 89
7.5. Connected Cars 90
7.5.1. Geolocalisation 90
7.5.2. Data forwarding 90
7.5.3. Internet of vehicles 90
7.6. Robots and drones 91
7.7. MAC mechanisms 91
7.8. RFID 92
7.9. Smart Cities 92
8. Bilateral Contracts and Grants with Industry **93**
9. Partnerships and Cooperations **93**
9.1. Regional Initiatives 93
9.1.1. StoreConnect 93
9.1.2. PIPA 93
9.2. National Initiatives 94
9.2.1. Inria Project Lab 94
9.2.2. ADT 94
9.2.3. Equipements d’Excellence 94
9.3. European Initiatives 95
9.4. International Initiatives 95
9.4.1. Inria International Labs 95

9.4.2. Inria International Partners 96
9.4.2.1. Declared Inria International Partners 96
9.4.2.2. Informal International Partners 96
9.4.3. Participation in Other International Programs 96
9.5. International Research Visitors 97
10. Dissemination **97**
10.1. Promoting Scientific Activities 97
10.1.1. Scientific Events Organisation 97
10.1.2. Scientific Events Selection 97
10.1.2.1. Chair of Conference Program Committees 97
10.1.2.2. Member of the Conference Program Committees 97
10.1.3. Journal 98
10.1.3.1. Guest editorial activities 98
10.1.3.2. Member of the Editorial Boards 98
10.1.4. Invited Talks 98
10.1.5. Leadership within the Scientific Community 99
10.1.6. Scientific Expertise 99
10.2. Teaching - Supervision - Juries 99
10.2.1. Teaching 99
10.2.2. Supervision 100
10.2.3. Juries 100
10.3. Popularization 101
11. Bibliography **101**

Project-Team FUN

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

Keywords:

Computer Science and Digital Science:
A1.2.1. - Dynamic reconfiguration
A1.2.3. - Routing
A1.2.4. - QoS, performance evaluation
A1.2.5. - Internet of things
A1.2.6. - Sensor networks
A1.2.7. - Cyber-physical systems
A1.4. - Ubiquitous Systems
A5.10.6. - Swarm robotics

Other Research Topics and Application Domains:
B5.1. - Factory of the future
B5.6. - Robotic systems
B5.9. - Industrial maintenance
B6.4. - Internet of things
B7. - Transport and logistics
B8. - Smart Cities and Territories

1. Team, Visitors, External Collaborators

Research Scientists
Nathalie Mitton [Team leader, Inria, Senior Researcher, HDR]
Valeria Loscri [Inria, Researcher, HDR]

Faculty Member
Antoine Gallais [Univ de Strasbourg, Associate Professor, HDR]

Technical Staff
Matthieu Berthome [Inria, Engineers]
Antonio Costanzo [Inria, Engineers]
Rehan Malak [Inria, Engineers, from April 2018]
Ibrahim Amadou [Inria, Engineers, from Jan 2018]
Julien Vandaele [Inria]

PhD Students
Abdoul Aziz Mbacke [Inria, PhD Students, Until Oct 2018]
Brandon Foubert [Inria, PhD Students, from Sept 2018]
Jad Nassar [HEI, PhD Students, Until Oct 2018]
Joao Batista Pinto Neto [UFRJ, PhD Students, Until March 2018]

Post-Doctoral Fellows
Allan Blanchard [Inria, Post-Doctoral Fellows]
Farouk Mezghani [Inria, Post-Doctoral Fellows, until August 2018]
Anjalalaina Jean Cristanel Razafimandimby [Univ des sciences et technologies de Lille]

Visiting Scientists

Pranvera Kortoci [Aalto University, from Mar 2018 until Apr 2018]
Noura Mares [Univ. Sfax, from Apr 2018 until Jul 2018]
Morgan O Kennedy [Stellenbosch University, from Jul 2018 until Sep 2018]

Administrative Assistant

Anne Rejl [Inria]

2. Overall Objectives

2.1. Overall Objectives

Context.

The Internet of Things [52] 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 [61].

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 FUN research group. Some of them are very efficient for routing [54], [51], discovering neighborhood [59], [58], scheduling activity and coverage [56], localizing [62], [50], 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 [67], [55] and at the MAC layer level [60]. 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.

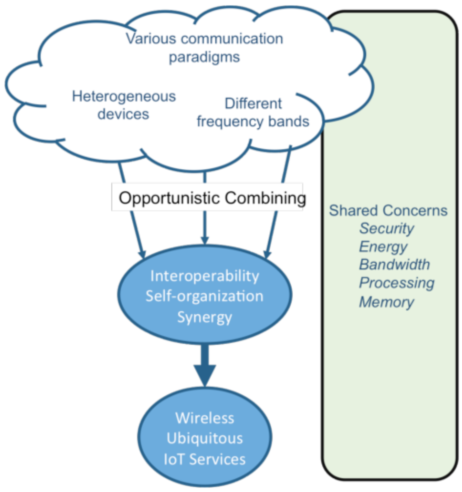

Figure 1. FUN's overal goal.

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 orthogonal 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 [68]. 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 [68], [25].

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 [63], [49]. On the same idea, when specific features are available like controlled mobility, we will exploit it to improve connectivity or coverage quality like in [57], [66].

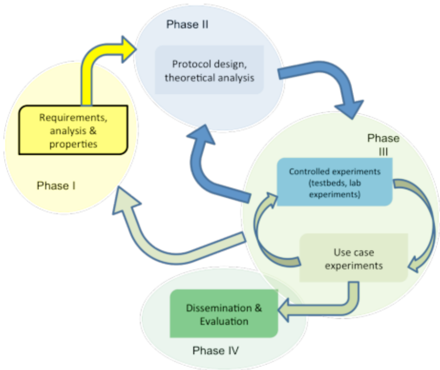

Figure 2. Methodology to be applied in FUN.

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 [64]. 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 [39] 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 [14], [23].

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 [53]. 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, environmental or animal monitoring, understanding and protection. To illustrate this, we may refer to the use cases addressed by our AgriNet project which goals is to monitor vineyards and potatoes fields with smart communicating devices to fight against water waste.

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 can play a key role in logistics and traceability by relying on the use of sensors or RFID technologies as implemented in our STORECONNECT 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).

5. Highlights of the Year

5.1. Highlights of the Year

- Valeria Loscrí has been elevated to the IEEE Senior Membership degree.

- Valeria Loscrí and Nathalie Mitton highlighted as Women in Computer Science 2018 by Elsevier https://www.elsevier.com/physical-sciences-and-engineering/computer-science/journals/women-in-computer-science.

5.1.1. Awards

Paper [65] has been awarded Top Paper from 2017 in Transactions on Emerging Telecommunications Technologies, as selected by the Editorial Board. See https://onlinelibrary.wiley.com/page/journal/21613915/homepage/ett_best_paper_awards.htm.

VITALOS, issued from the FP7 VITAL project from which the FUN team is one of the main contributors, has been awarded as one of the 50 most transformative smart projects: https://spring.smartcitiesconnect.org/Smart50Awards/.

6. New Software and Platforms

6.1. AspireRFID ALE

FUNCTIONAL DESCRIPTION: AspireRFID middleware is a modular OW2 open source RFID middleware. It is compliant with EPC Global standards. This new module integrates the modifications of the new standard release, including new RP and LLRP definitions and fixing bugs. This module has been implemented in the framework of the MIAOU project.

- Participants: Ibrahim Amadou, Julien Vandaële, Nathalie Mitton and Rim Driss
- Contact: Nathalie Mitton

6.2. ETINODE-CONTIKI-PORT

FUNCTIONAL DESCRIPTION: Contiki is an open source embedded OS for Internet of Things (IoT). It is light and portable to different hardware architectures. It embeds communication stacks for IoT Il embarque aussi des piles de communication pour l'internet des objets. This driver allows the running of Contiki OS over Etinode-MSP430. The code dalso allows the use of radio chip and embedded sensors. This module has been implemented in the framework of the ETIPOPS project.

- Participants: Nathalie Mitton, Roudy Dagher and Salvatore Guzzo Bonifacio
- Contact: Salvatore Guzzo Bonifacio

6.3. ETINODE-DRIVERS

FUNCTIONAL DESCRIPTION: These drivers for Etinode-MSP430 control the different embedded sensors and hardware components available on an Etinode-MSP430 node such as gyroscope, accelerometer and barometric sensor. This module has been implemented in the framework of the ETIPOPS project.

- Participants: Nathalie Mitton, Roudy Dagher and Salvatore Guzzo Bonifacio
- Contact: Salvatore Guzzo Bonifacio

6.4. EVe-TCF

Embedded Verifier for Transitive Control Flow
KEYWORDS: Control Flow - JavaCard - Embedded systems - Embedded - Security - Code analysis

FUNCTIONAL DESCRIPTION: Verification of transitive control flow policies on JavaCard 2.x bytecode. Control flow policies expressed using a DSL language are embedded in JavaCard packages (CAP files) using EVe-TCF convert tool. Control flow policies are then statically verified on-device at loading-time thanks to an embedded verifier (designed for smart cards in EVe-TCF). EVe-TCF (Embedded Verifier for Transitive Control Flow) also contains an off-device (i.e. PC tool) to simulate on-device loading process of JavaCard 2.x platforms with GlobalPlatform 2.x installed.

- Participants: Arnaud Fontaine and Isablle Simplot Ryl
- Contact: Nathalie Mitton

6.5. GOLIATH

Generic Optimized LIghtweight communication stack for Ambient TecHnologies
KEYWORDS: WSN - WSN430
FUNCTIONAL DESCRIPTION: GOLIATH (Generic Optimized LIghtweight communication stack for Ambient TecHnologies) is a full protocol stack for wireless sensor networks. This module has been implemented in the framework of the ETIPOPS project.

- Participants: David Simplot Ryl, Fadila Khadar, Nathalie Mitton and Salvatore Guzzo Bonifacio
- Contact: Nathalie Mitton
- URL: https://gforge.inria.fr/projects/goliath/

6.6. IoT-LAB robots

KEYWORDS: Internet of things - Robotics
FUNCTIONAL DESCRIPTION: IoT-LAB robots is an embedded robot controler on a Turtlebot2 providing the IoT-LAB node mobility functionnality

- Partner: Université de Strasbourg
- Contact: Julien Vandaële
- URL: https://github.com/iot-lab/

6.7. T-SCAN

KEYWORDS: Rfid - RFID Middleware
FUNCTIONAL DESCRIPTION: T-Scan is an interface ensuring the translation from a SGTIN tag format to an ONS hostname format according to the EPCGlobal standards. It allows the sending of a DNS request to look up the EPC-IS aides to which the product belongs in order to access the data relative to that product. This module has been implemented in the framework of the TRACAVERRE project.

- Participants: Gabriele Sabatino and Nathalie Mitton
- Contact: Gabriele Sabatino

6.8. FIT IoT-Lab

Participants: Nathalie Mitton [correspondant], Julien Vandaele, Matthieu Berthome.

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. Performance Evaluation, Security, Safety and Verification

Participants: Antoine Gallais, Nathalie Mitton, Allan Blanchard.

7.1.1. Performance Evaluation and validation methodology

Envisioned communication densities in Internet of Things applications are increasing continuously. Because these wireless devices are often battery powered, we need specific energy efficient (low-power) solutions. Moreover, these smart objects use low-cost hardware with possibly weak links, leading to a lossy network. Once deployed, these low-power lossy networks (LLNs) are intended to collect the expected measurements, handle transient faults, topology changes, etc. Consequently, validation and verification during the protocol development are a matter of prime importance. A large range of theoretical or practical tools are available for performance evaluation. A theoretical analysis may demonstrate that the performance guarantees are respected, while simulations or experiments aim on estimating the behavior of a set of protocols within real-world scenarios. In [16], we review the various parameters that should be taken into account during such a performance evaluation. Our primary purpose is to provide a tutorial that specifies guidelines for conducting performance evaluation campaigns of network protocols in LLNs. We detail the general approach adopted in order to evaluate the performance of layer 2 and 3 protocols in LLNs. Furthermore, we also specify the methodology that should be adopted during the performance evaluation, while reviewing the numerous models and tools that are available to the research community.

7.1.2. Correlated failures

Current practices of fault-tolerant network design ignore the fact that most network infrastructure faults are localized or spatially correlated (i.e., confined to geo-graphic regions). Network operators require new tools to mitigate the impact of such region-based faults on their infrastructures. Utilizing the support from the U.S. Department of Defense, and by consolidating a wide range of theories and solutions developed in the last few years, [14] designs RAPTOR, an advanced Network Planning and Management Tool that facilitates the design and provisioning of robust and resilient networks.The tool provides multi-faceted network design, evaluation, and simulation capabilities for network planners. Future extensions of the tool currently being worked upon not only expand the tool's capabilities, but also extend these capabilities to heterogeneous interdependent networks such as communication, power, water, and satellite networks.

7.1.3. Contiki verification

Internet of Things (IoT) applications are becoming increasingly critical and require formal verification. Our recent work presented formal verification of the linked list module of Contiki, an OS for IoT. It relies on a parallel view of a linked list via a companion ghost array and uses an inductive predicate to link both views. In this work, a few interactively proved lemmas allow for the automatic verification of the list functions specifications, expressed in the acsl specification language and proved with the Frama-C/Wp tool. In a broader verification context, especially as long as the whole system is not yet formally verified, it would be very useful to use runtime verification , in particular, to test client modules that use the list module. It is not possible with the current specifications, which include an inductive predicate and axiomatically defined functions. In

[27], an early-idea paper we show how to define a provably equivalent non-inductive predicate and a provably equivalent non-axiomatic function that belong to the executable subset e-acsl of acsl and can be transformed into executable C code. Finally, we propose an extension of Frama-C to handle both axiomatic specifications for deductive verification and executable specifications for runtime verification.

In [23], [47], we target Contiki, a widely used open-source OS for IoT, and present a verification case study of one of its most critical modules: that of linked lists. Its API and list representation differ from the classical linked list implementations, and are particularly challenging for deductive verification. The proposed verification technique relies on a parallel view of a list through a companion ghost array. This approach makes it possible to perform most proofs automatically using the Frama-C/WP tool, only a small number of auxiliary lemmas being proved interactively in the Coq proof assistant. We present an elegant segment-based reasoning over the companion array developed for the proof. Finally, we validate the proposed specification by proving a few functions manipulating lists.

With the wide expansion of multiprocessor architectures, the analysis and reasoning for programs under weak memory models has become an important concern. [13] presents MMFilter, an original constraint solver for generating program behaviors respecting a particular memory model. It is implemented in Prolog using CHR (Constraint Handling Rules). The CHR formalism provides a convenient generic solution for specifying memory models. It benefits from the existing optimized implementations of CHR and can be easily extended to new models. We present MMFilter design, illustrate the encoding of memory model constraints in CHR and discuss the benefits and limitations of the proposed technique.

7.2. Alternative communication paradigms

Participants: Antonio Costanzo, Valeria Loscri.

Nowadays, the always growing of connected objects and the strong demand to downsizing the devices in order to make the Internet of Things (IoT) paradigm more pervasive and ubiquitous, has motivated academic and industry people to investigate from one side mechanisms able to adapt quickly to the rapid external changes and to the quality of Services (QoS) parameters defined by the users and imposed by the adoption of new services and from another side, the investigation of portion of spectrum that have not been considered till this moment such as Terahertz band.

Nowadays, the always growing of connected objects and the strong demand to downsizing the devices in order to make the Internet of Things (IoT) paradigm more pervasive and ubiquitous, has motivated academic and industry people to investigate from one side mechanisms able to adapt quickly to the rapid external changes and to the quality of Services (QoS) parameters defined by the users and imposed by the adoption of new services and from another side, the investigation of portion of spectrum that have not been considered till this moment such as Terahertz band. In order to be able to realize a paradigm shift towards the Internet of Everything concept, a downsizing of devices is imposed allowing new applications as *in-vivo* diagnosis and monitoring. In order to be effective at this level it is imperative to analyze the new context, by highlighting the unique features to concretely realize the IoE paradigm. In this context, we have studied quantum particles called phonons, quasi-particles derived from vibrations of atoms in solids. Phonons have been envisaged as enabler of information transfer and their special characteristics have been exploited in [17]. Phonons have been also considered for a quantum channel in [26]. Another interesting approach for enabling the nano communication paradigm is represented by molecular communication. In particular, a main issue that is important to face is the coexistence between an artificial molecular communication and a biological system as explained in [40]. Alternative communication paradigms have attracted a lot of attention in the last a few years, not only by academic researchers but also by industry. Research on optical communication and in particular the possible exploitation of Visible Light communication with a twofold objective, to illuminate and to communicate has been object of an increasing interest. In this directions, we have proposed context-aware VLC systems in[39], [38] and [24]. The context is different in respect to the “traditional” wireless communication, since the external environment can change fastly and abruptly. Based on this primary observation, our main objective is to make theVLC system aware of the external noise and try to make it as robust as possible in respect of it.

7.3. Self-Organization

Participants: Antoine Gallais, Nathalie Mitton, Valeria Loscri, Farouk Mezghani, Anjalalaina Jean Cristanel Razafimandimby.

7.3.1. Stable parent selection

The Industrial Internet of Things consists in the use of low power lossy networks to enable next industrial applications. To work properly, the network has to provide strict guarantees concerning the delay and the reliability. IEEE 802.15.4-TSCH proposes time synchronized and slow channel hopping medium access control to cope with these requirements. It relies on a strict schedule of the transmissions, spread over orthogonal radio channels, to set up a resilient wireless infrastructure. A routing protocol (e.g. RPL) has then to construct energy-efficient routes on top of this link-layer topology (as investigated in the 6TiSCH IETF working group). Most of existing solutions rely on tree-based topologies, where each node has to select one or multiple parents to forward its traffic to the destination. Unfortunately, the links to the routing parents exhibit time-varying characteristics, due to e.g. obstacles, and external interference, thus leading to oscillations and increased required control of the routing topology. Moreover, the network has to provision enough resources (i.e., time, channel) to cope with those variations, while still being reactive to node/link failures. We investigated the stability of 6TiSCH networks, and especially the impact on routing parent selection. We identified moments of instability due to oscillations in the radio conditions caused by external interference and obstacles, in two indoor testbeds with different channel conditions. We identified the causes of instabilities, and proposed solutions for each of the layers in the 6TiSCH stack. First, at the MAC layer, we demonstrated that a rearrangement of shared cells in the slotframe reduces the probability of collisions for control packets, paving the way to a faster negotiation during topology reconfigurations. Next, we eased the schedule consistency management between two nodes (renegociated from scratch in the current standard, upon detection of a schedule inconsistency). Finally, at the routing layer, we exploited the existing correlation between the broadcast packet reception rate and the unicast link quality to create a two-step parent selection that favors stable parents. We finally obtained a network that converged faster and that reacted accurately during moments of instabilities. Results are available in [46], [42].

7.3.2. Bayesian communications

The amount of data that are generated in IoT devices is huge and the most of time data are highly correlated, by making useless the forwarding of all the raw data generated. Bearing that in mind, we have designated and implemented an effective mechanism to reduce the amount of data sent in the network in [45]. Results are encouraging since there is a size effect of less interfering in the communication system with an important impact on battery consumption for wireless devices that are energy constrained.

7.3.3. Multi-technology self-organization

Opportunistic communications present a promising solution for disaster network recovery in emergency situations such as hurricanes, earthquakes, and floods, where infrastructure might be destroyed. Some recent works in the literature have proposed opportunistic-based disaster recovery solutions, but they have omitted the consideration of mobile devices that come with different network technologies and various initial energy levels. [19], [30] present COPE, an energy-aware Cooperative OPportunistic alErt diffusion scheme for trapped survivors to use during disaster scenarios to report their position and ease their rescue operation. It aims to maintain mobile devices functional for as long as possible for maximum network coverage until reaching proximate rescuers. COPE deals with mobile devices that come with an assortment of networks and aims to perform systematic network interface selection. Furthermore, it considers mobile devices with various energy levels and allows low-energy nodes to hold their charge for longer time with the support of high-energy nodes. A proof-of-concept implementation has been performed to study the doability and efficiency of COPE, and to highlight the lessons learned. Following-up with these results, we performed several experimentations and could benchmark smartphone performances with regards to their multi-communications interfaces. Testing experiments have been carried out to measure the performance of smartphones in terms of energy consumption, clock synchronization and transmission range. We believe that such experimental results can support technological choices for rescue operations but also for many other applications relying on smartphone performances. Results are available in [30].

7.3.4. Heterogeneous Self-organizing (smart) Things

In the panorama of the Internet of Things, one main important issue is the management of heterogeneous objects, that need to communicate in order to exchange information and to interact in order to be able to synergically accomplish complex tasks and for providing services to final users. In this context, the thesis [10] has tried to face the main challenges related to complex heterogeneous systems, where objects are able to self-organize to each other and are equipped with some kind of intelligence in order to dynamically react to the environment changes. Several tools have been exploited ranging from artificial neural networks to genetic algorithms and different solutions have been proposed to make these systems dynamic and responding to the $*$self properties.

7.4. Smart Grids

Participants: Nathalie Mitton, Jad Nassar.

The Smart Grid (SG) aims to transform the current electric grid into a “smarter” network where the integration of renewable energy resources, energy efficiency and fault tolerance are the main benefits. This is done by interconnecting every energy source, storage point or central control point with connected devices, where heterogeneous SG applications and signaling messages will have different requirements in terms of reliability, latency and priority. Hence, data routing and prioritization are the main challenges in such networks.

So far, RPL (Routing Protocol for Low-Power and Lossy networks) protocol is widely used on Smart Grids for distributing commands over the grid. RPL assures traffic differentiation at the network layer in wireless sensor networks through the logical subdivision of the network in multiple instances, each one relying on a specific Objective Function. However, RPL is not optimized for Smart Grids, as its main objective functions and their associated metric does not allow Quality of Service differentiation.

In order to overcome this, we propose $OFQS$ an objective function [20] with a multi-objective metric that considers the delay and the remaining energy in the battery nodes alongside with the dynamic quality of the communication links. Our function automatically adapts to the number of instances (traffic classes) providing a Quality of Service differentiation based on the different Smart Grid applications requirements. We tested our approach on a real sensor testbed. The experimental results show that our proposal provides a lower packet delivery latency and a higher packet delivery ratio while extending the lifetime of the network compared to solutions in the literature.

The management of communication is an issue in WSN-based Smart Grid: billions of messages with different sizes and priorities are sent across the network. Data aggregation is a potential solution to reduce loads on the communication links, thus achieving a better utilization of the wireless channel and reducing energy consumption. On the other hand, SG applications require different Quality of Service (QoS) priorities. Delays caused by data aggregation must then be controlled in order to achieve a proper communication. In [33], [34], we propose a work in progress, that consists of a QoS efficient data aggregation algorithm with two aggregation functions for the different traffics in a SG network. We expect to reduce the energy consumption while respecting the data delivery delays for the different SG applications.

In order to reduce the amount of data sent over the network, and thus reduce energy consumption, data prediction is another potent solution of data reduction. It consists on predicting the values sensed by sensor nodes within certain error threshold, and resides both at the sensors and at the sink. The raw data is sent only if the desired accuracy is not satisfied, thereby reducing data transmission. We focus on time series estimation with Least Mean Square (LMS) for data prediction in WSN, in a Smart Grid context, where several applications with different data types and Quality of Service (QoS) requirements will exist on the same network. LMS proved its simplicity and robustness for a wide variety of applications, but the parameters selection (step size and filter length) can directly affect its global performance, choosing the right ones is then crucial. Having no clear and robust method on how to optimize these parameters for a variety of applications, we propose in [44] a modification of the original LMS that consists of training the filter for a certain time with the data itself in order to customize the aforementioned parameters. We consider different types of real data traces for the photo voltaic cells monitoring. Our simulation results provide a better data prediction while minimizing the mean square error compared to an existing solution in literature.

All these solutions have also been detailed in [12].

7.5. Connected Cars

Participants: Nathalie Mitton, Valeria Loscri, Joao Batista Pinto Neto.

7.5.1. Geolocalisation

Connected car technology promises to drastically reduce the number of accidents involving vehicles. Nevertheless, this technology requires the vehicle precise location to work. The adoption of Global Positioning System (GPS) as a navigation device imposes limitations to geolocation information under non-line-of-sight conditions. [22] introduces the Time Series Dead Reckoning System (TedriS) as a solution for dead reckoning navigation when the GPS fails. TedriS uses Time Series Regression Models (TSRM) and the data from the rear wheel speed sensor of the vehicle to estimate the absolute position. The process to estimate the position is carried out in two phases: training and predicting. In the training phase, a novel technique applies TSRM and stores the relationship between the GPS and the rear wheel speed data; then in the predicting phase, this relationship is used. We analyze TedriS using traces collected at the campus of Federal University of Rio de Janeiro (UFRJ), Brazil, and with indoor experiments with a robot. Results show an accuracy compatible with dead-reckoning navigation state-of-art systems.

7.5.2. Data forwarding

Intelligent inter-vehicle communication is a key research field in the context of vehicular networks that applies in real-life applications (e.g., management of accidents, intelligent fuel consumption, smart traffic jams, etc.). Considering different roles of nodes based on their "social aptitude" to relay information could provide a social component in the vehicular structure that can be useful in getting a clear prediction of the topological evolution in time and space proving to be very effective in managing intelligent data forwarding. In [36], we characterize a vehicular network as a graph using the link layer connectivity level and we classify nodes on the basis of specific attributes characterizing their "social aptitude" to forward data. Two forwarding approaches are presented, based on different socialites that allow to (i) select the most social node (i.e., a social hub) or (ii) choose among various social nodes.

7.5.3. Internet of vehicles

Internet, in its most recent evolution, is going to be the playground where a multitude of heterogeneous interconnected "things" autonomously exchange information to accomplish some tasks or to provide a service. Recently, the idea of giving to those smart devices the capability to organize themselves according to a social structure, gave birth to the so-called paradigm of the Social Internet of Things. The expected benefits of SIoT range from the enhanced effectiveness, scalability and speed of the navigability of the network of interconnected objects, to the provision of a level of trustworthiness that can be established by averaging the social relationships among things that are "friends". Bearing in mind the beneficial effects of social components in IoT, we consider a social structure in a vehicular context i.e., Social Internet of Vehicles (SIoV). In SIoV, smart vehicles build social relationships with other social objects they might come into contact, with the intent of creating an overlay social network to be exploited for information search and dissemination for vehicular applications. In [43], we aim to investigate the social behavior of vehicles in SIoV and how it is affected by mobility patterns. Specifically, through the analysis of simulated traffic traces, we distinguish friendly and acquaintance vehicles based on the encounter time and connection maintenance.

7.6. Robots and drones

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

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. In this context, we propose a Neuro-Dominating Set algorithm (NDS) [21] to efficiently deploy a team of mobile wireless robots in an IoRT scenario, in order to reach a desired inter-robot distance, while maintaining global connectivity in the whole network. We use the term Neuro-Dominating Set to describe our approach, since it is inspired by both neural network and dominating set principles. With NDS algorithm, a robot adopts different behaviors according whether it is a dominating or a dominated robot. Our main goal is to show and demonstrate the beneficial effect of using different behaviors in the IoRT concept. The obtained results show that the proposed method outperforms an existing related technique (i.e., the Virtual Angular Force approach) and the neural network based approach presented in our previous work. As an objective, we aim to decrease the overall traveled distance and keep a low energy consumption level, while maintaining network connectivity and an acceptable convergence 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. This is what we propose in [32] with the objective is to maximize the exploration of the given surface. To ensure that 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.

When a disaster strikes, the telecommunications infrastructure gets damaged making rescue operations more challenging. Connecting first responders through flying base stations (i.e. drone mounted LTE (Long-Term Evolution) femtocell base station) presents a promising alternative to support infrastructure failure during disasters. The drone can travel the area and communicate with ground mobile devices, such as smartphones, and serves as flying data link to share information between survivors and rescuers. Problem statement. We would like to submit the following open problem to the community. Given the position of the ground mobile devices to serve, the problem presented here is about the dynamic drone path planning. As the drone autonomy is very limited and due to the high cost of drone mounted base station, the goal of this problem is to determine the best energy-efficient and minimum-time path to travel the area as fast as possible while still remaining in range of each survivor long enough to assure full servicing. This is the problem stated in [31].

7.7. MAC mechanisms

Participant: Nathalie Mitton.

In the era of the Internet of Things (IoT), the number of connected devices is growing dramatically. Often, connected objects use Industrial, Scientific and Medical (ISM) radio bands for communication. These kinds of bands are available without license, which facilitates development and implementation of new connected objects. However, it also leads to an increased level of interference in these bands. Interferences not only negatively affect the Quality of Service, but also cause energy losses, which is especially unfavorable for the energy constrained Wireless Sensor Networks (WSN). In [25], we develop an explicit formula of outage probability in a distributed wireless sensor network (WSN), assuming the MAC layer protocol being a slotted-ALOHA. And adopting a Markovian approach, we develop a model that analyses the performance of the slotted-ALOHA in order to improve these performances, in particular, by adding a preliminary stage of channel reservation, we show that this modification is important to have a high performance distributed wireless sensor network.

Several wild animal species are endangered by poaching. As a solution, deploying wireless sensors on animals able to send regular messages and also alert messages has been envisaged recently by several authorities and foundations. In that context, we have proposed WildMAC [35], a multichannel, multihop wireless communication protocol for these specific wireless sensor networks that have to collect data from unknown large areas with different QoS requirements. WildMAC is a TDMA based MAC protocol that leverages long range communication properties to propose an efficient data collection mean. Its performance evaluation shows it meets QoS requirements. To size the different parameters of WildMAC, we relied on the results of the study of [25].

7.8. RFID

Participants: Nathalie Mitton, Abdoul Aziz Mbacke, Ibrahim Amadou.

While RFID technology is gaining increased attention from industrial community deploying different RFID-based applications, it still suffers from reading collisions. As such, many proposals were made by the scientific community to try and alleviate that issue using different techniques either centralized or distributed, monochannel or multichannels, TDMA or CSMA. However, the wide range of solutions and their diversity make it hard to have a clear and fair overview of the different works. [18] surveys the most relevant and recent known state-of-the-art anti-collision for RFID protocols. It provides a classification and performance evaluation taking into consideration different criteria as well as a guide to choose the best protocol for given applications depending on their constraints or requirements but also in regard to their deployment environments.

Among all these approaches, [29], [28] propose new reader anti-collision schemes and data-priority aware data collection in a multi-hop RFID data collection protocol. [28] examines the implementation of two applications: for industrial IoT and for smart cities, respectively. Both applications, in regards to their requirements and configuration, challenge the operation of a RFID sensing solution combined with a dynamic wireless data gathering over multihops. They require the use of both mobile and fixed readers to cover the extent of deployment area and a quick retrieval of tag information. We propose a distributed crosslayer solution for improving the efficiency of the RFID system in terms of collision and throughput but also its proficiency in terms of tag information routing towards one or multiple sinks. Simulation results show that we can achieve high level of throughput while maintaining a low level of collision and a fairness of reader medium access above 95% in situations where readers can be fix and mobile, while tag information is routed with a data rate of 97% at worst and reliable delays for considered applications. [29] proposes cross-layer solutions meant for both scheduling of readers' activity to avoid collisions, and a multihop routing towards base stations, to gather read tag data. This routing is performed with a data priority aware mechanism allowing end-to-end delay reduction of urgent data packets delivery up to 13% faster compared to standard ones. Using fuzzy logic, we combine several observed metrics to reduce the load of forwarding nodes and improve latency as well as data rate. We validate our proposal running simulations on industrial and urban scenarios.

All these solutions have also been detailed in [11].

7.9. Smart Cities

Smart cities are a key factor in the consumption of materials and resources. As populations grow and resources become scarcer, the efficient usage of these limited goods becomes more important. Building on and integrating with a huge amount of data, the cities of the future are becoming a realization today. There are millions of sensors in place already, monitoring various things in metropolises. In the near future, these sensors will multiply until they can monitor everything from streetlights and trashcans to road conditions and energy consumption. In this context, effective strategies or solutions for refining data sets can play a key role. In [37], we propose a scheme in which passive RFID is shown as an interesting alternative and complement to WSN to alleviate the cost of some Smart City applications.

Also, in Smart Cities, crowd sensing may help to identify the current speed for each street, the congested areas, etc. In this context, map matching techniques are required to map a sequence of GPS waypoints into a set of streets on a common map. Unfortunately, most map matching approaches are probabilistic. In [41], we propose rather an unambiguous algorithm, able to identify all the possible paths that match a given sequence of waypoints. We need an unambiguous identification for each waypoints set. For instance, the actual speed should be assigned to the correct set of streets, without error. To identify all the possible streets, we construct the set of candidates iteratively. We identify all the edge candidates around each waypoint, and reconstruct all the possible sub-routes that connect them. We then verify a set of constraints, to eliminate impossible routes. The road segments common to all computed routes form an unambiguous match. We evaluate the matching ratio of our technique on real city maps (London, Paris and Luxembourg). We also validate our approach with a real GPS trace in Seattle.

In parallel, we proposed a MOOC in the framework of the IPL CityLab project (See Section 9.2.1), whose working documents are available online [48].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

- Sencrop
  Participants: Brandon Foubert, Nathalie Mitton [correspondant].

  This collaboration aims to develop a complete multi-technology bilateral wireless communication stack for agriculture sensor networks.
- Enedis and NooliTic
  Participants: Ibrahim Amadou, Nathalie Mitton [correspondant].

  This collaboration aims to investigate a novel localization approach based on wireless propagations. It is a tri-partite contract between our Inria team, the SME NooliTic and Enedis.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. StoreConnect

Participants: Nathalie Mitton, Valeria Loscri [correspondant], Antonio Costanzo, Ibrahim Amadou.

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. This project has yielded to several publications in 2018: [39], [38], [24], [40].

9.1.2. PIPA

Participants: Nathalie Mitton [correspondant], Farouk Mezghani.

Title: Partager de l'Info PArtout à bas coût

Type: Chercheur citoyen

Duration: Dec 2015 - Aug 2018

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. This project has yield to several publications in 2018: [31], [19], [30].

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 [48] and a set of publications [18], [29], [28], [37], [11].

9.2.2. ADT

9.2.2.1. Catimex

Participants: Matthieu Berthome, Nathalie Mitton [correspondant], Julien Vandaele.

Duration: September 2017 -June 2019

Coordinator: Inria FUN

The purpose of this project is to foster research transfer in IoT from ADT members to their industrial partners by widening experimental features and PoC realization. It is lead in closed partnership with Inria Chile and Université of Strasbourg.

9.2.3. Equipements d'Excellence

9.2.3.1. FIT

Participants: Nathalie Mitton [correspondant], Julien Vandaele, Matthieu Berthome.

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.

9.3. European Initiatives

9.3.1. H2020 Projects

9.3.1.1. VESSEDIA

Participants: Rehan Malak, Nathalie Mitton, Allan Blanchard [correspondant].

- Title: VERIFICATION ENGINEERING OF SAFETY AND SECURITY CRITICAL DYNAMIC INDUSTRIAL APPLICATIONS
- Program: 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, organizations 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 standardization 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. This project yields to set of publications in 2018: [23], [47], [27].

9.4. International Initiatives

9.4.1. Inria International Labs

9.4.1.1. Agrinet

Participants: Abdoul Aziz Mbacke, Brandon Foubert, Valeria Loscri, Anjalalaina Jean Cristanel Razafimandimby, Nathalie Mitton [correspondant].

Agrinet

Title: Agrinet

International Partner (Institution - Laboratory - Researcher):

Type: LIRIMA Associate team

Duration: 2017-2020

See also: https://team.inria.fr/agrinet/

The current drought and limited water resources in many parts of Southern Africa and beyond, already have a significant impact on agriculture and hence, food production. Sustainable food security depends upon proper plant and crop management respectful of soils and natural re- sources, such as water. This includes very important South African farming areas, such as the Western Cape and Northern Cape. In France, agriculture is also hugely important. Not just nationally, but also in Europe. The system proposed can be applied to a variety of crops. The economic- and social consequences are profound and any contribution towards more efficient farming within increasingly onerous natural constraints, should be a priority. To address these constraints, we propose to develop a flexible, rapidly deployable, biological/agricultural data acquisition platform and associated machine learning algorithms to create advanced agricultural monitoring and management techniques, to improve crop management and use of natural resources. The project also addresses an industry with very high socioeconomic impact.

Publications issued from that project in 2018 are: [25], [45], [35].

9.4.2. Inria International Partners

9.4.2.1. 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.

9.4.2.2. 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 and joint publications.

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 2018 [14].

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 2018 [17], [21], [26], [36], [43], [45].

9.4.3. Participation in Other International Programs

9.4.3.1. CROMO

Participants: Valeria Loscri, Joao Batista Pinto Neto, Nathalie Mitton [correspondant].

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 [22].

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Several researchers have visited our group in 2018, mainly from our partner universities but not only:

- Gentian, Jakllari, ENSEEITH, France, January 2018
- Georgios Papadopoulos, IMT Atlantique, France, January 2018
- Bruno Quoitin, University of Mons, Belgium, January and March 2018
- Sebastien Bindel, Université de Haute Alsace, France, June 2018
- Karen Miranda Campos, Metropolitan Autonomous University Lerma Campus, Mexico, October 2018
- Zied Chtourou, University of Sfax, Tunisia, October 2018
- Fabrice Théoleyre, University of Strasbourg, October 2018
- Fabrice Valois, INSA Lyon, October 2018
- Miguel Elias Campista, Federal University of Rio de Janeiro (UFRJ), Brazil, December 2018
- Pranvera Kortoci, Aalto University, from Mar 2018 until Apr 2018
- Noura Mares Univ. Sfax, from Apr 2018 until Jul 2018
- Morgan O Kennedy, Stellenbosch University, from Jul 2018 until Sep 2018

9.5.1.1. Research Stays Abroad

- Jad Nassar visited Metropolitan Autonomous University Cuajimalpa Campus, Mexico in January-February 2018

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 demo/poster chair of infocom 2019
- Valeria Loscri is general co-chair of City-Wide Pervasive Environment CoWPER, in conjunction with IEEE SECON 2018, and of a Special Session on Mobile Social Networks in PIMRC 2018
- Antoine Gallais was the co-chair of IIoT day 2018

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- Antoine Gallais is TPC co-chair of CoRes 2019
- Valeria Loscri is TPC co-chair of IEEE ACM Nanocom 2019 and of IEEE ACM SmartComp in conjunction with Mobihoc 2018
- Nathalie Mitton is TPC co-chair of ICC 2019 and 2018

10.1.2.2. Member of the Conference Program Committees

- Valeria loscri is/was member in the Technical Program Committee (TPC) in IFIP/IEEE NTMS 2018, Globecom 2018-2019, CORES 2018, VTC 2018-2019, GIoTS 2018-2019, WiMob 2018, CCNC 2018-2019, ICC 2019
- Nathalie Mitton is/was in the Technical Program Committee (TPC) of Percom 2019, WCNC 2019, DCOSS 2019, CORES 2018, Infocom workshop 2018, VTC 2018, globecom 2019&2018, GIIS 2018, CSCN2018, ICC 2018, coconet 2018, adhocnow 2018
- Antoine Gallais was/is in the Technical Program Committee (TPC) of IEEE Globecom'19 and 2018, IEEE ICNC'19 and 2018, IEEE GIIS'18, IARIA INNOV'18, IEEE COMNETSAT'18, IARIA EMERGING'18, ADHOC-NOW'18, ICST Adhocnets'18
- Allan Blanchard was/is in the Technical Program Committee (TPC) of 4PAD 2018

10.1.3. Journal

10.1.3.1. Guest editorial activities

- Nathalie Mitton was (co-)editor of Sensor Special Issues of MDPI Sensors on QoS in WSN and of MDPI Future Internet WSN in Smart Agriculture
- Valeria Loscri was editor of IEEE Access Special Section on Protocols for Nanocommunication Networks, of IEEE Access Special Section on Body Area Networks and of a Special Issue on IEEE Internet of Things Journal on Recent Advances on Social Internet of Vehicles

10.1.3.2. 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
- Nathalie Mitton is editorial board member of MDPI Future Internet since 2018
- Valeria Loscri is editorial board member of IEEE Transactions on Nanobioscience journal since 2017
- Valeria Loscri is editorial board member of Elsevier Computer Networks journal since 2016
- Valeria Loscri is editorial board member of Robotics Software Design and Engineering of the International Journal of Advanced Robotic Systems since 2016
- Valeria Loscri is editorial board member of Elsevier Journal of Networks and Computer Applications (JNCA) journal since 2016

10.1.4. Invited Talks

- Nathalie Mitton was invited speaker at Workshop "Digital Agriculture in Africa", in Montpellier, in April 2018
- Nathalie Mitton was invited speaker at CORES conference in Roscoff, France, in May 2018
- Nathalie Mitton was invited speaker at WFCS conference in Imperia, Italy, in June 2018
- Nathalie Mitton was invited speaker at Nantes University, France, in October 2018
- Antoine Gallais was invited speaker at Innov'School Network of Things (ISNOT), Gammarth, Tunisia in April 2018
- Allan Blanchard gave an invited tutorial at ZINC 2018, Novi Sad, Serbia in June 2018
- Allan Blanchard gave an invited tutorial at HPCS 2018, Orleans, France, in July 2018

- Valeria Loscri was invited as panelist in the panel "Inspirations from wireless networks to biology" in the 3rd EU FET CIRCLE Workshop, Ghent April 2018
- Valeria Loscri was invited as panelist in the Panel New Trends in Communication and at the Doctoral Colloquium at ACM/IEEE NanoCom 2018
- Valeria Loscri has been appointed as Faculty Mentor in N2 Women meeting at ACM Mobihoc 2018
- Valeria Loscri was invited to attend the Seed meeting on Dynamical Modeling and Simulation for Molecular Communication that was held at the Embassy of France in London
- Valeria Loscri was invited speaker at the LiFi seminar in Vélizy, September 2018

10.1.5. Leadership within the Scientific Community

- Nathalie Mitton is a member of the Steering Committee of the GDR Rescom
- Valeria Loscri is a member of Social Network Technical Committee
- Valeria Loscri is a member of Emerging Technologies Initiatives for Molecular, Biological and Multi-Scale Communications (ETI-MBMC)
- Valeria Loscri is a member of the Quantum Communications & Information Technology Emerging Technical Subcommittee (QCIT)
- Valeria Loscri is a member of the "Research Group on IoT Communications and Networking Infrastructure" at ComSoc Communities

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 and as an evaluator for Chilean National Science and Technology Commission (CONICYT - Chile) FONDECYT Initiation into Research 2018 projects. She is also member of the scientific committees of the competitiveness cluster of MATIKEN and for CITC (International Contactless Technologies Center). Finally, Nathalie Mitton is a member of the HCERES visiting committee for the LISIS laboratory.
- Valeria Loscri is Scientific European Responsible for Inria Lille - Nord Europe. She is reviewer in the context of ERC Consolidator Grant. She is reviewer for Equipes Associées. She has been scientific reviewer of TOP grant for senior researchers in the context of Netherlands Organisation for Scientific Research (NWO) program.
- Valeria Loscri has been external expert reviewer of proposal for grant competition at the Center for Excellence in Applied Computational Science and Engineering (CEACSE), UTC (USA).

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, in November 2018

SPOC, Nathalie Mitton,EIT Digital "Technological challenges of participatory Smart cities", 5-week by the IPL CityLab@Inria team in November 2018

Remote curse, Nathalie Mitton, Internet of things, 5-week + face to face week in May 2018

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

Master: Nathalie Mitton, Introduction to Internet of Things, 8h CM Ecole Centrale de Lille, France

Master: Ibrahim Amadou, Introduction to Internet of Things, 8h TP Ecole Centrale de Lille, France

Master: Ibrahim Amadou and Abdoul Aziz Mbacke, Introduction to RFID, 16h IMT Rennes, France

Master: Ibrahim Amadou, Wireless sensor networks, 16h eqTD (Master ROC), IMT, France

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

Licence: Jean Razafimandimby, Algorithms and Programming, 30h eqTD, Université Lille 1, France

10.2.2. Supervision

PhD defended on October 18th 2018: 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 defended on October 12th 2018: Jad Nassar, Ubiquitous networks for smart grids, Université Lille 1, 2015-2018, Nathalie Mitton and Nicolas Gouvy (HEI)

PhD in progress: Brandon Foubert, Communication sans fil Polymorphe pour l'Agriculture Connectée, Université Lille 1, 2018-2021, Nathalie Mitton

PhD in progress: Rodrigo Teles, Virtualizing Heterogeneous Wireless Networks with SDN for the 5G, Université Strasbourg, 2016-2019, Antoine Gallais and Fabrice Théoleyre

PhD in progress: M. Amine Falek, Optimization algorithms for personalized and dynamic multi-modal itinerary planning, Université Strasbourg, 2016-2029, Antoine Gallais, Cristel Pelsser and Fabrice Theoleyre

PhD in progress: Loïc Miller, Secure work-flow access control, Université Strasbourg, 2018-2021, Antoine Gallais, Pascal Merindol and Cristel Pelsser

10.2.3. Juries

- PhD/HDR committees:
  - Antoine Gallais is/was members of the following PhD thesis committees:
    - François Lemercier, IMT Atlantique, Rennes, France, November 2018
    - Solomon Petrus le Roux, Stellenbosch University, South Africa, December 2018
    - Hugo Chelle, Universite´ de Toulouse, France, December 2018
  - Valeria Loscri is/was members of the following PhD thesis committees:
    - Alexis Duque, INSA Lyon, France, October 2018
    - Óscar Alvear, Universidade Politecnica Valencia, Spain, July 2018
    - Sabrine Aroua, Université de La Rochelle, France, July 2018
    - Mohamed Abdelkrim, Université Paris Est Créteil, December 2018
  - Nathalie Mitton is/was member of the following PhD thesis committees:
    - Patrick Olivier Kamgueu, Univ. de Yaounde - Univ. de Lorraine, January 2018
    - Hermes Pimenta de Moraes Junior, UTC (présidente), May 2018
    - Hamadoun Tall, Clermont Auvergne, (reviewer) May 2018
    - Marwan Ghanem, UPMC, September 2018
    - Wafa Badreddine, UPMC, (reviewer) November 2018
    - Yosra Zguira, INSA Lyon (reviewer), November 2018
    - Solomon Petrus le Roux (reviewer), Stellenbosch University, December 2018
    - Fadhlallah Baklouti, Université Bretagne Sud (reviewer), February 2019
    - M Naas, Université de Bretagne Occidentale (reviewer), February 2019
  - Nathalie Mitton was a member of the HDR defense committees

    * Valeria Loscri, Université Lille 1, March. 2018
    * Fen Zhou, Université d'Avignon, Sept. 2018

- Researcher selection committees:
  - Nathalie Mitton was a member of the Professor competition selection committee at Ecole des Mines Nancy, Université Nice-Sophia Antipolis and Clermont-Ferrand
  - Nathalie Mitton was a member of the Inria chaire committee for Supelec Rennes
  - Nathalie Mitton was a member of the Assistant Professor (MdC) at Valenciennes
- Nathalie Mitton is/was reviewer of the following PhD follow-up committees:
  - Rodrigue Domgua Rodriguez, INSA Lyon
  - Yosra Zguira, INSA Lyon
  - Vaseileios Kotsiou, Université de Strasbourg
- Antoine Gallais is/was reviewer of the following PhD follow-up committees:
  - Tomas Lagos Jenschke, IMT Atlantique

10.3. Popularization

- Matthieu Berthome, Jad Nassar, Ibrahim Amadou, Antonio Costanzo and Antoine Gallais gave some talks in high schools for the "Fête de la Science" in October 2018
- Nathalie gave an interview for France Info on connected objects

11. Bibliography

Major publications by the team in recent years

[1] R. T. HERMETO, A. GALLAIS, F. THEOLEYRE.*On the (over)-Reactions and the Stability of a 6TiSCH Network in an Indoor Environment*, in "21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems (MSWIM '18)", Montreal, Canada, October 2018 [*DOI :* 10.1145/3242102.3242104], https://hal.archives-ouvertes.fr/hal-01886716

[2] K. KRITSIS, G. PAPADOPOULOS, A. GALLAIS, P. CHATZIMISIOS, F. THEOLEYRE.*A Tutorial on Performance Evaluation and Validation Methodology for Low-Power and Lossy Networks*, in "Communications Surveys and Tutorials, IEEE Communications Society", July 2018, vol. 20, n^{o} 3, p. 1799 - 1825 [*DOI :* 10.1109/COMST.2018.2820810], https://hal.archives-ouvertes.fr/hal-01886690

[3] V. LOSCRI, B. D. UNLUTURK, A. MARIA VEGNI.*A Molecular Optical Channel Model based on Phonon-Assisted Energy Transfer Phenomenon*, in "IEEE Transactions on Communications", August 2018, https://hal.inria.fr/hal-01864323

[4] A. A. MBACKÉ, N. MITTON, H. RIVANO.*A survey of RFID readers anticollision protocols*, in "IEEE Journal of Radio Frequency Identification", March 2018, vol. 2, n^{o} 1, 11 [*DOI :* 10.1109/JRFID.2018.2828094], https://hal.inria.fr/hal-01767311

[5] J. NASSAR, M. BERTHOMÉ, J. DUBRULLE, N. GOUVY, N. MITTON, B. QUOITIN.*Multiple Instances QoS Routing In RPL: Application To Smart Grids*, in "Sensors", August 2018, vol. 18, n^{o} 8 [*DOI :* 10.3390/s18082472], https://hal.inria.fr/hal-01851713

[6] J. B. Pinto Neto, N. Mitton, M. E. M. Campista, L. H. M. K. Costa. *Dead Reckoning Using Time Series Regression Models*, in "MobiHoc 2018 - 4th ACM MobiHoc Workshop on Experiences with the Design and Implementation of Smart Objects", Los Angeles, United States, June 2018, vol. 18 [*DOI :* 10.1145/3213299.3213305], https://hal.inria.fr/hal-01798550

[7] C. Razafimandimby, V. Loscri, A. M. Vegni, A. Benslimane. *Neuro-Dominating Set Scheme for a Fast and Efficient Robot Deployment in Internet of Robotic Things*, in "Ad Hoc Networks", August 2018, https://hal.inria.fr/hal-01864325

[8] R. Teles Hermeto, A. Gallais, F. Theoleyre. *Passive Link Quality Estimation for Accurate and Stable Parent Selection in Dense 6TiSCH Networks*, in "International Conference on Embedded Wireless Systems and Networks (EWSN)", Madrid, Spain, February 2018, p. 114–125, https://hal.archives-ouvertes.fr/hal-01886778

[9] A. M. Vegni, V. Loscri, P. Manzoni. *Data Forwarding Techniques Based on Graph Theory Metrics in Vehicular Social Networks*, in "IEEE PIMRC 2018 - 29th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications", Bologna, Italy, September 2018, https://hal.inria.fr/hal-01826237

Publications of the year

Doctoral Dissertations and Habilitation Theses

[10] V. Loscri. *Toward Interoperability of Heterogeneous Self-organizing (smart) Things*, University of Lille 1, March 2018, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01743527

[11] A. A. Mbacké. *Collection and multi-hop forwarding of RFID data for the monitoring of urban infrastructures*, Université de Lille, October 2018, https://hal.inria.fr/tel-01901740

[12] J. Nassar. *Ubiquitous Networks for Smart Grids*, Universite des Sciences et Technologies de Lille, October 2018, https://hal.inria.fr/tel-01908825

Articles in International Peer-Reviewed Journal

[13] A. Blanchard, N. Kosmatov, F. Loulergue. *MMFilter: A CHR-Based Solver for Generation of Executions under Weak Memory Models*, in "Computer Languages, Systems and Structures", September 2018, https://hal.inria.fr/hal-01777123

[14] A. Das, A. Sen, C. Qiao, N. Ghani, N. Mitton. *RAPTOR: a network tool for mitigating the impactof spatially correlated failures in infrastructure networks*, in "Annals of Telecommunications - annales des télécommunications", February 2018, vol. 73, n° 1, p. 153-164, https://hal.inria.fr/hal-01584867

[15] M. Egan, V. Loscri, T. Q. Duong, M. D. Renzo. *Strategies for Coexistence in Molecular Communication*, in "IEEE Transactions on NanoBioscience", 2019, https://hal.archives-ouvertes.fr/hal-01928205

[16] K. Kritsis, G. Papadopoulos, A. Gallais, P. Chatzimisios, F. Theoleyre. *A Tutorial on Performance Evaluation and Validation Methodology for Low-Power and Lossy Networks*, in "Communications Surveys and Tutorials, IEEE Communications Society", July 2018, vol. 20, n° 3, p. 1799 - 1825 [*DOI :* 10.1109/COMST.2018.2820810], https://hal.archives-ouvertes.fr/hal-01886690

[17] V. LOSCRI, B. D. UNLUTURK, A. MARIA VEGNI.*A Molecular Optical Channel Model based on Phonon-Assisted Energy Transfer Phenomenon*, in "IEEE Transactions on Communications", August 2018, https://hal.inria.fr/hal-01864323

[18] A. A. MBACKÉ, N. MITTON, H. RIVANO.*A survey of RFID readers anticollision protocols*, in "IEEE Journal of Radio Frequency Identification", March 2018, vol. 2, n^{o} 1, 11 [*DOI :* 10.1109/JRFID.2018.2828094], https://hal.inria.fr/hal-01767311

[19] F. MEZGHANI, N. MITTON.*Opportunistic Disaster Recovery*, in "Internet Technology Letters", January 2018, p. 1-5 [*DOI :* 10.1002/ITL2.29], https://hal.archives-ouvertes.fr/hal-01691285

[20] J. NASSAR, M. BERTHOMÉ, J. DUBRULLE, N. GOUVY, N. MITTON, B. QUOITIN.*Multiple Instances QoS Routing In RPL: Application To Smart Grids*, in "Sensors", August 2018, vol. 18, n^{o} 8 [*DOI :* 10.3390/S18082472], https://hal.inria.fr/hal-01851713

[21] C. RAZAFIMANDIMBY, V. LOSCRI, A. M. VEGNI, A. BENSLIMANE.*Neuro-Dominating Set Scheme for a Fast and Efficient Robot Deployment in Internet of Robotic Things*, in "Ad Hoc Networks", August 2018, https://hal.inria.fr/hal-01864325

Invited Conferences

[22] J. B. PINTO NETO, N. MITTON, M. E. M. CAMPISTA, L. H. M. K. COSTA.*Dead Reckoning Using Time Series Regression Models*, in "MobiHoc 2018 - 4th ACM MobiHoc Workshop on Experiences with the Design and Implementation of Smart Objects", Los Angeles, United States, June 2018, vol. 18 [*DOI :* 10.1145/3213299.3213305], https://hal.inria.fr/hal-01798550

International Conferences with Proceedings

[23] A. BLANCHARD, N. KOSMATOV, F. LOULERGUE.*Ghosts for Lists: A Critical Module of Contiki Verified in Frama-C*, in "Tenth NASA Formal Methods Symposium - NFM 2018", Newport News, United States, April 2018, https://hal.inria.fr/hal-01720401

[24] A. COSTANZO, V. LOSCRI.*Demo: A Context Aware Algorithm for an Adaptive Visible Light Communication System*, in "EWSN 2018 - International Conference on Embedded Wireless Systems and Networks", Madrid, Spain, February 2018, https://hal.inria.fr/hal-01686565

[25] M. EL AMINE SEDDIK, V. TOLDOV, L. CLAVIER, N. MITTON.*From Outage Probability to ALOHA MAC Layer Performance Analysis in Distributed WSNs*, in "WCNC 2018 - IEEE Wireless Communications and Networking Conference", Barcelona, Spain, April 2018, https://hal.inria.fr/hal-01677687

[26] V. LOSCRI, A. M. VEGNI.*Error Probability Derivation in a Phonon-based Quantum Channel*, in "QCIT'18 - Workshop on Quantum Communications and Information Technology", Abu Dhabi, France, December 2018, https://hal.inria.fr/hal-01877153

[27] F. LOULERGUE, A. BLANCHARD, N. KOSMATOV.*Ghosts for Lists: from Axiomatic to Executable Specifications*, in "TAP 2018 - 12th International Conference on Tests and Proofs", Toulouse, France, June 2018, https://hal.inria.fr/hal-01811922

[28] A. A. MBACKÉ, N. MITTON, H. RIVANO.*DACAR: Distributed & Adaptable Crosslayer Anticollision and Routing protocol for RFID*, in "AdHoc-Now 2018 - 17th International Conference on Ad Hoc Networks

and Wireless", St Malo, France, Springer, September 2018, p. 226-238 [*DOI :* 10.1007/978-3-030-00247-3_21], https://hal.inria.fr/hal-01819767

[29] A. A. MBACKÉ, N. MITTON, H. RIVANO.*Using Fuzzy Logic for data priority aware collection in RFID sensing wireless networks*, in "PIMRC 2018 - IEEE International Symposium on Personal, Indoor and Mobile Radio Communications", Bologne, Italy, September 2018, p. 1-5, https://hal.inria.fr/hal-01830721

[30] F. MEZGHANI, N. MITTON.*Benchmarking smartphone performances for cooperative disaster alert diffusion*, in "AdHoc-Now 2018 - 17th International Conference on Ad Hoc Networks and Wireless", Saint Malo, France, September 2018, https://hal.inria.fr/hal-01826642

[31] F. MEZGHANI, N. MITTON.*Open problem: Energy-and time-efficient dynamic drone path planning for post-disaster network servicing*, in "ODS 2018 - International Conference on Optimization and Decision Science", Taormina, Italy, September 2018, https://hal.inria.fr/hal-01826657

[32] S. MICHEL, L. BROTCORNE, D. C. CATTARUZZA, N. MITTON, F. SEMET.*A heuristic approach for the computation of individual trajectories of a fleet of robots under connectivity constraints*, in "Congrès annuel de la société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF)", Lorient, France, February 2018, https://hal.inria.fr/hal-01704705

[33] J. NASSAR, N. GOUVY, N. MITTON.*QoS-compliant Data Aggregation for Smart Grids*, in "ENERGY 2018 - The Eighth International Conference on Smart Grids, Green Communications and IT Energy-aware Technologies", Nice, France, May 2018, https://hal.inria.fr/hal-01744647

[34] J. NASSAR, K. MIRANDA, N. GOUVY, N. MITTON.*Prédiction de données différenciée pour les Smart Grids* , in "Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01785425

[35] V. TOLDOV, L. CLAVIER, N. MITTON.*Multi-channel Distributed MAC protocol for WSN-based wildlife monitoring*, in "WiMob 2018 - 14th International Conference on Wireless and Mobile Computing, Networking and Communications", Limassol, Cyprus, October 2018, https://hal.inria.fr/hal-01866809

[36] A. M. VEGNI, V. LOSCRI, P. MANZONI.*Data Forwarding Techniques Based on Graph Theory Metrics in Vehicular Social Networks*, in "IEEE PIMRC 2018 - 29th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications", Bologna, Italy, September 2018, https://hal.inria.fr/hal-01826237

National Conferences with Proceeding

[37] A. A. MBACKÉ, N. MITTON, H. RIVANO.*Une vision plus durable et fiable de la ville intelligente grâce à la RFID*, in "CORES 2018 - Rencontres Francophones sur la Conception de Protocoles, l'Évaluation de Performance et l'Expérimentation des Réseaux de Communication", Roscoff, France, May 2018, p. 1-4, https://hal.archives-ouvertes.fr/hal-01784572

Conferences without Proceedings

[38] A. COSTANZO, V. LOSCRI.*Algorithme de récupération de porteuse pourdes systèmes de communication de lumièrevisible adaptées au contexte*, in "ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01782098

[39] A. COSTANZO, V. LOSCRI, S. COSTANZO.*Adaptive Dual Color Visible Light Communication (VLC) System*, in "6th World Conference on Information Systems and Technologies", Napoli, Italy, March 2018, https://hal.inria.fr/hal-01687543

[40] M. EGAN, T. Q. DUONG, M. DI RENZO, J.-M. GORCE, I. NEVAT, V. LOSCRI.*Cognitive Molecular Communication*, in "2018 -3rd Workshop on Molecular Communications", Ghent, Belgium, April 2018, p. 1-2, https://hal.archives-ouvertes.fr/hal-01728766

[41] M. A. FALEK, C. PELSSER, A. GALLAIS, S. JULIEN, F. THEOLEYRE.*Unambiguous, Real-Time and Accurate Map Matching for Multiple Sensing Sources*, in "WiMob 2018 - 14th International Conference on Wireless and Mobile Computing, Networking and Communications", Limassol, Cyprus, October 2018, https://hal.archives-ouvertes.fr/hal-01886732

[42] R. T. HERMETO, A. GALLAIS, F. THEOLEYRE.*On the (over)-Reactions and the Stability of a 6TiSCH Network in an Indoor Environment*, in "MSWIM '18 - 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems", Montreal, Canada, October 2018 [*DOI :* 10.1145/3242102.3242104], https://hal.archives-ouvertes.fr/hal-01886716

[43] V. LOSCRI, G. RUGGERI, A. M. VEGNI, I. CRICELLI.*Social Structure Analysis in Internet of Vehicles*, in "CoWPER Workshop in conjunction with IEEE International Conference on Sensing, Communication and Networking", Hong-Kong, China, June 2018, https://hal.archives-ouvertes.fr/hal-01798609

[44] J. NASSAR, K. MIRANDA, N. GOUVY, N. MITTON.*Heterogeneous data reduction in WSN: Application to Smart Grids*, in "SMARTOBJECTS 2018 - 4th Workshop on Experiences with the Design and Implementation of Smart Objects", Los Angeles, California, United States, June 2018, https://hal.archives-ouvertes.fr/hal-01773861

[45] C. RAZAFIMANDIMBY, V. LOSCRI, A. M. VEGNI, A. NERI.*Une approche bayésienne pour une réduction efficace des données dans l'IdO*, in "ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01778326

[46] R. TELES HERMETO, A. GALLAIS, F. THEOLEYRE.*Passive Link Quality Estimation for Accurate and Stable Parent Selection in Dense 6TiSCH Networks*, in "EWSN 2018 - International Conference on Embedded Wireless Systems and Networks", Madrid, Spain, February 2018, p. 114–125, https://hal.archives-ouvertes.fr/hal-01886778

Other Publications

[47] A. BLANCHARD, N. KOSMATOV, F. LOULERGUE.*Des listes et leurs fantômes : vérification d'un module critique de Contiki avec FRAMA-C*, June 2018, Résumé d'article, https://hal.inria.fr/hal-01811932

[48] N. MITTON, V. ISSARNY, A. BAYLE.*MOOC Défis technologiques des villes intelligentes participatives - Session 3. Contributions aux activités de fin de semaine 2 : Des vélos connectés pour de meilleures pistes cyclables et de nombreux services*, March 2018, Contributions aux activités de fin de semaine 2 du MOOC Villes intelligentes. Document édité par l'équipe pédagogique du MOOC Villes Intelligentes sur la base des contributions des participants Session 3 - novembre 2017, https://hal.inria.fr/hal-01736139

References in notes

[49] 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

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

[51] 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

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

[53] 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

[54] 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

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

[56] 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

[57] 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

[58] 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

[59] 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

[60] A. A. MBACKÉ, N. MITTON, H. RIVANO.*RFID Anticollision in Dense Mobile Environments*, in "WCNC 2017 - IEEE Wireless Communications and Networking Conference", San Francisco, United States, March 2017, https://hal.inria.fr/hal-01418170

[61] 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

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

[63] 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

[64] R. PETROLO, V. LOSCRI, N. MITTON, E. HERRMANN.*Adaptive Filtering as a Service for Smart City Applications*, in "14th IEEE International Conference on Networking, Sensing and Control (ICNSC)", Cosenza, Italy, May 2017, https://hal.inria.fr/hal-01482715

[65] 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

[66] 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

[67] 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

[68] V. TOLDOV.*Adaptive MAC layer for interference limited WSN*, Université Lille 1 Sciences et technologies, January 2017, https://hal.inria.fr/tel-01493094

Team GAIA

Geometry, Algebra, Informatics, Applications

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
Algorithmics, Computer Algebra and Cryptology

Table of contents

1. **Team, Visitors, External Collaborators** **111**
2. **Overall Objectives** **111**
   2.1. Algebraic and geometric studies of functional systems 111
   2.2. Effective algebraic theories and their implementations in computer algebra systems 112
   2.3. A rich interplay between computer algebra and control theory 113
   2.4. Main objectives of the GAIA team 113
3. **Research Program** **114**
   3.1. Effective algebra 114
   3.1.1. Polydisc Nullstellensatz & effective version of a theorem of Deligne 114
   3.1.2. Effective version of Spencer's theory of formal integrability of PD systems 114
   3.1.3. Rings of integro-differential operators & integro-differential algebra 114
   3.1.3.1. Linear systems of integro-differential equations 114
   3.1.3.2. Nonlinear systems of integro-differential equations 115
   3.2. Computer algebra 115
   3.2.1. Efficient algorithms for the study of singularities of algebraic curve and its applications 115
   3.2.2. Differential algebra 116
   3.2.3. A Maple package and a C/C++ open source library for integro-differential algebra 116
   3.3. Applications to control theory and signal processing 116
   3.3.1. Robust stability analysis and stabilization problems for functional systems 116
   3.3.1.1. Computation of Lyapunov functions for homogeneous dynamical systems 116
   3.3.1.2. Robust stability analysis of differential time-delay systems 117
   3.3.1.3. Stabilization problems for functional systems 117
   3.3.2. Parameter estimation for linear & nonlinear functional systems 117
   3.3.2.1. Linear functional systems 117
   3.3.2.2. Nonlinear functional systems 118
4. **Application Domains** **118**
   4.1. Adaptive & parametric robust control – collaboration with Safran Electronics & Defense 118
   4.2. Self calibration problem & Gear fault diagnosis – collaboration with Safran Tech 119
   4.2.1. Self calibration problem 119
   4.2.2. Gear fault diagnosis 120
   4.3. Applications of the parameter estimation problem to multidisciplinary domains – collaboration with an INSERM team (Rouen University) 120
5. **Highlights of the Year** **120**
6. **New Software and Platforms** **120**
   6.1. ADHOMFI 120
   6.2. Platforms 121
7. **New Results** **121**
   7.1. Regular (differential) chains 121
   7.2. Systems of integro-differential equations 121
   7.3. Certified non-conservative tests for the structural stability of discrete multidimensional systems 121
   7.4. Using symbolic computation to solve algebraic Riccati equations arising in invariant filtering 121
   7.5. Parametric sub-optimal H_∞ controllers for an optro-mechanical system 122
   7.6. A symbolic approach for signal demodulation and application to gearbox vibration analysis 122
   7.7. Curve analysis for the stability of time-delay systems 122
8. **Bilateral Contracts and Grants with Industry** **122**
   8.1. Safran Electronics & Defense 122
   8.2. Ellcie Healthy 122
9. **Partnerships and Cooperations** **123**

9.1. National Initiatives 123
9.2. European Initiatives 123
9.3. International Initiatives 123
9.4. International Research Visitors 124
10. Dissemination **124**
10.1. Promoting Scientific Activities 124
10.1.1. Scientific Events Organisation 124
10.1.1.1. General Chair, Scientific Chair 124
10.1.1.2. Member of the Organizing Committees 124
10.1.2. Scientific Events Selection 124
10.1.2.1. Member of the Conference Program Committees 124
10.1.2.2. Reviewer 124
10.1.3. Journal 124
10.1.3.1. Member of the Editorial Boards 124
10.1.3.2. Reviewer - Reviewing Activities 124
10.1.4. Invited Talks 125
10.1.5. Leadership within the Scientific Community 125
10.1.6. Scientific Expertise 125
10.1.7. Research Administration 125
10.2. Teaching - Supervision - Juries 125
10.2.1. Teaching 125
10.2.2. Supervision 125
10.2.3. Juries 125
10.3. Popularization 126
10.3.1. Internal or external Inria responsibilities 126
10.3.2. Education 126
10.3.3. Internal action 126
11. Bibliography **126**

Team GAIA

Creation of the Team: 2018 January 01, end of the Team: 2018 December 31

Keywords:

Computer Science and Digital Science:
A6.4. - Automatic control
A8.3. - Geometry, Topology
A8.4. - Computer Algebra

Other Research Topics and Application Domains:
B6.6. - Embedded systems
B9.5.1. - Computer science
B9.5.2. - Mathematics

1. Team, Visitors, External Collaborators

Research Scientists
Alban Quadrat [Team leader, Inria, Senior Researcher, HDR]
Rosane Ushirobira [Inria, Researcher, HDR]

Faculty Members
Francois Boulier [Univ des sciences et technologies de Lille, Professor, from Mar 2018]
Francois Lemaire [Univ des sciences et technologies de Lille, Associate Professor, from Mar 2018]
Adrien Poteaux [Univ des sciences et technologies de Lille, Associate Professor]
Michele Triestino [Univ de Bourgogne, Associate Professor, from Nov 2018]

PhD Students
Guillaume Rance [Safran Electronics & Defense, until Sep 2018]
Grace Younes [Inria, from Oct 2018]

Post-Doctoral Fellows
Cyrille Chenavier [Inria, from Oct 2018]
Yacine Bouzidi [Inria]

Visiting Scientists
Marc Moreno Maza [Univ Western Ontario, from Sep 2018 until Oct 2018]
Fredrik Olsson [Univ Uppsala, Nov 2018]

Administrative Assistant
Corinne Jamroz [Inria]

2. Overall Objectives

2.1. Algebraic and geometric studies of functional systems

Systems of functional equations or simply *functional systems* are systems whose unknowns are functions, such as systems of ordinary (OD) or partial differential (PD) equations, of differential time-delay equations, of difference equations, of integro-differential equations, etc. [34], [35]. Functional systems play a fundamental role in the mathematical modeling of physical phenomena studied in natural science such as physics, or in engineering sciences such as mathematical systems theory control theory, signal processing, etc. [34], [35]. Numerical aspects of functional systems, especially OD and PD systems, have largely been studied in applied mathematics due to the importance of numerical simulation issues.

Complementary approaches, based on algebraic and differential or algebraic geometric methods, are usually upstream or help the numerical simulation of systems of functional systems. These methods tackle questions and problems such as algebraic preconditioning, elimination and simplification, completion to formal integrability or involution, computation of integrability conditions or compatibility conditions, index reduction, reduction of variables, choice of adapted coordinate systems based on symmetries, computation of first integrals of motion, conservation laws, and Lax pairs, study of Liouville integrability or of the asymptotic behavior of solutions at a singularity, etc. For more details, see [36], [41], [51], [67], [75], [76], [81], [85], [101], [104], [109] and the references therein.

Let us state a few interests of an algebraic approach for the study of functional systems:

- Algebraic methods are clearly suitable for an algorithmic study, and thus for the development of efficient algorithms implementable in computer algebra systems.
- It can be used to finely study the behavior of the solutions of a system with respect to unfixed model parameters – which is usually a difficult numerical issue. Moreover, the boundaries of the zones in the parameter space over which the behavior of the solution changes can be algorithmically characterized by means of algebraic methods, which yields a safe use of numeric methods in each regular zones (symbolic-numeric methods).
- The existence of closed-form solutions can highly simplify certain problems studied in applications by avoiding the use of time-consuming optimization problems, and thus fits well with nowadays real-time applications.

The GAIA team aims to develop algebraic and geometric methods for the study of functional systems.

2.2. Effective algebraic theories and their implementations in computer algebra systems

Although not yet very popular in applied mathematics, algebraic and differential geometric approaches of functional systems have lengthly been studied in foundamental mathematics. We can state a few names of mathematical theories such as (differential) Galois theory, Lie groups, exterior differential systems, differential algebra, algebraic analysis, etc. [36], [41], [51], [67], [75], [76], [81], [85], [101], [104], [109].

Over the past years, some of these algebraic theories for the study of functional systems have been investigated in the computer algebra community within an algorithmic viewpoint, mostly driven by applications to engineering sciences such as mathematical systems theory and control theory.

Gröbner or Janet bases [40], [102] for noncommutative polynomial rings of functional operators or differential elimination techniques for differential systems [45], [46], [72], based on differential algebra [101], [76], are remarkable examples of those effective algebraic methods. They are nowadays implemented in standard computer algebra systems (e.g. `Maple`, `Mathematica`, `Magma`).

These effective algebraic approaches also form the algorithmic "engines" at the basis of the first developments of effective versions of modern algebraic theories (algebraic geometry, differential algebra, module theory and homological algebra over certain noncommutative polynomial rings of functional operators, algebraic analysis, etc.).

The above-mentioned results are just the tip of the iceberg and much more effort must be made in the future for making effective larger parts of fundamental mathematics and making them largely available in standard computer algebra systems. This "democratization" process towards the accessibility of fundamental mathematics is important, for instance, for educational issues where computers can be used to teach them to students, scientists of other communities and engineerings, and for learning by doing and computing. Further developing effective mathematics, making them accessible to a larger audience through dedicated software, and demonstrating them through interesting engineering problems are at the core of the GAIA team. The latter engineering problems are major sources of motivation for the development of effective algebraic theories and their implementations in computer algebra systems as explained in the next section.

2.3. A rich interplay between computer algebra and control theory

A major source of motivation for the development of the effective study of algebraic theories is represented by control theory issues. Indeed, certain problems studied in control theory can be better understood and finely studied by means of algebraic or geometric structures and techniques. The rich interplay between algebra, computer algebra, and control theory has a long history.

The first main paper on Gröbner bases [50] written by their creators, Buchberger, was published in Bose's book [42] on control theory of multidimensional systems since they play a fundamental role in this theory. They were the first main applications of Gröbner bases outside the field of algebraic geometry and they still play a fundamental role in multidimensional systems theory [43].

The differential algebra approach to nonlinear control theory [58], [59], [63], [84] was a major motivation for the effective study of differential algebra [101], [76] (differential elimination theory, triangular sets, regular chains, etc.) [45], [46], [72] and its implementations in `Maple`. Within this effective differential algebra approach to nonlinear control systems, observability, identifiability, parameter estimation, invertibility, differential flatness, etc., have received appealing and checkable algebraic characterizations.

Linear control theory [73] and multidimensional systems theory [42], [43] have recently been profoundly developed due to the so-called behavior approach [80], [83] and the module approach [64], [84]. Based on ideas of *algebraic analysis* [75], system properties of those systems (e.g. controllability, parametrizability, differential flatness) are intrinsically characterized by means of properties of certain algebraic structures (namely finitely presented left modules over noncommutative polynomial rings of functional operators). To effectively check the latter properties, the development of effective versions of two important algebraic theories, namely module theory [79] and homological algebra [103], had to be iniated based on functional elimination techniques (i.e. Gröbner or Janet basis techniques for noncommutative polynomial rings) [7], [8] (see also [91]). Dedicated packages, written in `Maple`, `Mathematica` and `GAP`, are now available.

The GAIA team wants to further develop its expertise in this direction by considering new classes of functional systems (e.g. differential varying/distributed delay systems, ordinary integro-differential systems) interesting in control theory and in signal processing.

2.4. Main objectives of the GAIA team

The first goal of the GAIA team is to study classes of *functional systems* which are interesting in practice (e.g. differential systems, differential constant/varying/distributed delay systems, ordinary integro-differential systems) by means of *algebraic* and *geometric* methods (algebraic analysis, algebraic/differential/noncommutative geometry, etc.), *computer algebra* (e.g. algorithmic, symbolic and symbolic-numeric methods, librairies), and *mathematical systems theory*. The systems to be investigated can be linear, nonlinear, continuous, discrete, or originated from real life applications.

The second goal of the GAIA team is to study important problems coming from:

- *control theory* (e.g. parametric robust control, stability and stabilization of multidimensional systems or of differential constant/distributed/time-varying delay systems)
- *signal processing* (e.g. parameter estimation problem, metric multidimensional unfolding, autocalibration)
- *multidisciplinary domains* (e.g. marine bivalves behavior, human-machine interaction, chemical reaction networks, ionic activities in neuroscience)

The third goal of the GAIA team is to develop (`Maple`, `Mathematica`, `C/C++`) *packages and librairies* dedicated to functional systems and to their applications, and in parallel, eventual *industry transfer* (e.g. *Safran Electronics & Defense*, *Safran Tech*, *Maplesoft*).

3. Research Program

3.1. Effective algebra

To develop a computational study of problems coming from control theory, signal processing, and multi-disciplinary domains, parts of algebraic theories must be studied within an effective approach: methods and theoretical results must be made algorithmic based on computer algebra techniques appropriated for efficient implementations in computer algebra systems.

3.1.1. Polydisc Nullstellensatz & effective version of a theorem of Deligne

The works on stability and stabilization problems of multidimensional systems, developed in the former ANR MSDOS (2014–2018), have shown the importance for developing an effective version of the module theory over the ring of rational functions without poles in the closed unit polydisc of $\mathbb{C}^n$ [90], [47]. The stabilizability (resp. the existence of a doubly coprime factorization) of a multidimensional system is related to a module-theoretical property (projectivity, resp. freeness) that has to be algorithmically verified prior to compute stabilizing controllers (resp. the standard Youla-Kučera parametrization of all stabilizing controllers). Based on the works [89], [90], in [47], we have recently proved that the stabilizability condition is related to the development of an algorithmic proof of the so-called *Polydisc Nullstellensatz* [49], a natural extension of Hilbert's *Nullstellensatz* for the above-mentioned ring (see e.g. [40]). In addition, the existence of a doubly coprime factorization is related to a theorem obtained by (the Fields medalist) Deligne with a non-constructive proof [90]. This theorem can be seen as an extension of the famous Quillen-Suslin theorem (Serre's conjecture) [77]. Based on our experience of the first implementation of the Quillen-Suslin theorem in the computer algebra system (`Maple`) [62], we aim to develop this effective framework as well as a dedicated `Maple` package.

3.1.2. Effective version of Spencer's theory of formal integrability of PD systems

A differential geometric counterpart of differential algebra and differential elimination theory [101], [76] is the so-called Spencer's theory of formal integrability and involutive PD systems [85], [104]. For linear PD systems, this theory can be seen as an intrinsic approach to Janet or Gröbner bases for noncommutatibe polynomial rings of PD operators. No complete algorithmic study of Spencer's theory has been developed yet. We aim to develop it as well as to implement it. The understanding of the connections between the different differential elimination theories (Janet [102] or Gröbner bases [40], Thomas decomposition [102], differential algebra [101], [76], Spencer's theory [85], exterior differential systems [51], etc.) will also be investigated. On a longer term, applications of Spencer's theory to Lie pseudogroups and their applications in mathematical physics (e.g. variational formulations based on Lie (pseudo)groups) [86] will be investigated and implemented.

3.1.3. Rings of integro-differential operators & integro-differential algebra

The main contribution of this axis is the development of effective elimination theories for both linear and nonlinear systems of integro-differential equations.

3.1.3.1. Linear systems of integro-differential equations

The rings of integro-differential operators are more complex than the purely differential case [96], [97] due to the existence of zero-divisors or the fact of having a *coherent ring* instead of a *Noetherian ring* [39]. We want to develop an algorithmic study of these rings. Following the direction initiated in [95] for the computation of zero divisors, we first want to develop algorithms for the computation of left/right kernels and left/right/generalized inverses of matrices with entries in such rings, and to use them to develop a module-theoretic approach to linear systems of integro-differential equations. Following [95], standard questions addressed within the computer algebra community such as the computation of rational/exponential/hyperexponential/etc. solutions will also be addressed. Moreover, famous Stafford's results [105], algorithmically studied in [96], [97] for rings of PD operators, are known to still hold for rings of integro-differential operators [39]. Their algorithmic extensions will be investigated and our corresponding implementation will be extended accordingly. Finally, following [93], [95], an algorithmically study of rings of integro-differential-delay operators will be further developed as well as their applications to the equivalence problem of differential constant/varying/distributed delay systems (e.g. Artstein's reduction, Fiagbedzi-Pearson's transformation) and their applications to control theory.

3.1.3.2. Nonlinear systems of integro-differential equations

Integro-differential algebra is an extension of Ritt-Kolchin's *differential algebra* [101], [76] that also includes integral operators. This extension is now attracting more attention in mathematics, computer algebras, and control theory. This new type of algebras will be algorithmically studied for integro-differential nonlinear systems. To do that, concepts such as integro-differential ideals and varieties have to be introduced and studied for developing an integro-differential elimination theory which extends the current differential elimination theory [45], [46], [72]. A `Maple` prototype will first be developed and then a `C` library when experience will be gained.

3.2. Computer algebra

We aim to further reinforce our expertise in the computer algebra aspects of functional systems and algebraic curves by attacking remaining technical obstacles and by considering new classes of functional systems, notably those coming from interesting applications in engineering sciences and particularly in control theory.

3.2.1. Efficient algorithms for the study of singularities of algebraic curve and its applications

On an algorithmic viewpoint, there are mainly four different approaches for the study of the singularities of plane algebraic curves. Let us shortly list them:

- The well-known *Newton-Puiseux* algorithm, initiated by Cramer and Puiseux, which follows an idea due to Newton. This approach has successively been improved in [61], [88], [9], [10].
- The *Extended Hensel Construction*, developed in [74] and recently improved in [82].
- The work [68] concentrates on the factorisation of polynomials defined over valued fields (the local study of a plane algebraic curve enters in this approach).
- The work [33] introduces the concept of an*approximate root*.

The first two methods are based on Puiseux series computations. They use techniques which are equivalent to the standard blowing-up of a singularity of an algebraic curve, which has the drawback to be bottlenecks in terms of complexity and practical efficiency. Nevertheless, the recent work [88] provides the best complexity currently known. The last two methods study singularities without computing Puiseux series. They both use the concept of an *extended evaluation*.

A recent very efficient algorithm for the factorisation of a univariate polynomial based on its Newton polygon has recently been obtained in [52]. The key ingredient of this algorithm is to work on the given polynomial and not after changes of variables as usually done in the literature.

To improve the complexity results of [68], [38], we want to combine the above different approaches using approximate roots and a generalization of the results of [52] to the context of [68].

The method proposed above is important in practice since it is based on well-known and efficiently implemented algorithms (mainly Newton iteration and gcd computations). Nevertheless, these algorithms involve technical difficulties on the computer science side: the main one is the need to improve the accuracy of computations due to truncations. These issues, including also run-time compilation, are well-studied in the `BPAS` library [0] based on specific data structures to deal with power series computation (a power series is represented by terms that have been computed and a program that enables to compute more terms when required).

Another part of the code development concerns issues on certified numerical computations for univariate polynomials (with algebraic coefficients) that will be used for the development of certified symbolic-numeric algorithms making effective the strategy proposed in [87].

[0] http://www.bpaslib.org/index.html

3.2.2. Differential algebra

A major bottleneck of computational differential algebra methods is the computation of greatest common divisors of multivariate commutative polynomials. Any algorithmic progress in this direction would highly improve the efficiency of differential algebra software such as, for instance, the C library BLAD [44]. Moreover, numerous computer algebra problems and related implementations could also highly profit from any success in this direction.

A major application of the effective differential algebra approach developed by GAIA's members [45], [46] is the possibility to reduce a nonlinear (implicit) differential system, particularly differential algebraic equations, to so-called regular chains of differentiation index 0, i.e. to systems which do not need differentiation of their equations to be rewritten as pure differential systems [69]. Based on our expertise on differential techniques, we want to study the consistent initialization problem and develop numerical integrators for nonlinear differential algebraic systems, and used them in the study of coupled algebraic and differential systems, interconnected systems, or networks [69].

3.2.3. A Maple package and a C/C++ open source library for integro-differential algebra

A package dedicated to nonlinear integro-differential equations will be developed in a Maple prototype and then in a standalone C/C++ open source library (as it was already done for the diffalg and DifferentialAlgebra packages). General purpose solvers such as Mapledsolve or pdesolve may call differential elimination methods for computing essential singular solutions of differential equations, for computing systems of polynomial differential equations admitting a given function as a solution, etc. On the long run, one may foresee enhanced general purpose solvers able to handle integro-differential equations processed through integro-differential elimination methods. It will rely on a sub-package dedicated to the problem of effectively handling integro-differential expressions.

These packages will rely on existing software such as BLAD, DifferentialAlgebra, and MABSys. It is worth pointing out that proof-of-concept methods are already available. See [29] and [2]. The collaboration with modelers will also enhance the software user-interface for a better usability.

The study of numerical integration of integro-differential equations (a necessary component of software dedicated to the parameter estimation problem) will also be further studied following the direction initiated in [29], leading to the Maple/C library BLINEIDE. This software has currently no widely available challenger.

The Maple prototype software dedicated to nonlinear integro-differential equations will also be implemented in a standalone C/C++ open source library, leading to software easier to integrate in modeling platforms such as OpenModelica. The GAIA team has quite some expertise in releasing software satisfying industrial standards: its C open source BLAD libraries, dedicated to differential elimination, are currently integrated in Maple and called through the Maple package DifferentialAlgebra. The Modelica programming language, which emphasizes programming with equations and permits to call external code, can integrate software dedicated to integro-differential equations developed in the GAIA team.

3.3. Applications to control theory and signal processing

3.3.1. Robust stability analysis and stabilization problems for functional systems

Our expertise in the computer algebra aspects to stability and stabilization problems for multidimensional systems and for differential constant/distributed/varying delay systems [5], [6], [48], [47] will further be developed.

3.3.1.1. Computation of Lyapunov functions for homogeneous dynamical systems

We shall investigate the possibility to develop a computer algebra package for the design of Lyapunov functions for homogeneous dynamical systems based on an effective study of the differential algebra of *generalized forms*, i.e. of Puiseux polynomials in signed powers [106], [107].

3.3.1.2. Robust stability analysis of differential time-delay systems

The symbolic-numeric study of the robust stability of a differential constant time-delay system with respect to the delay h, via the variation of the zero locus of the associated quasipolynomial $p(s, e^{-h\,s})$ [70] in the stability (resp. unstability) region $\mathbb{C}_- = \{s \in \mathbb{C} \mid \Re(s) < 0\}$ (resp. $\overline{\mathbb{C}_-} = \mathbb{C} \setminus \mathbb{C}_+$) of $\mathbb{C}$, initiated in [5], will be further developed. This problem is another motivation for the development of a fast numerical algorithm for the computation of Netwon-Puiseux series and its implementation in a `C` library. They will be used to study how the different branches of a quasipolynomial at a critical pair (namely $(h_{☆}, \omega_{☆}) \in \mathbb{R}_{>0} \times \mathbb{R}$ such that $p(i\,\omega_{☆}, e^{-i\,h\,\omega_{☆}}) = 0$) vary in $\mathbb{C}_-$ and in $\mathbb{C}_+$ with respect to h. See [5] and the references therein.

Moreover, in collaboration with Mouze (Centrale Lille, France), we want to develop an effective study of the ring $\mathcal{E} = \mathbb{R}(s)[e^{-h\,s}] \cap E$, where E is the ring of entire functions. The ring $\mathcal{E}$ plays an important role for differential time-delay systems [66], [78]. Effective computation of Smith normal forms for matrices with entries in $\mathcal{E}$ and its implementation in a symbolic-numeric package will have many applications for synthesis problems of differential time-delay systems.

3.3.1.3. Stabilization problems for functional systems

We want to use the results developed on the module-theoretic aspects of the ring of multivariate rational functions without poles in the closed unit polydisc of $\mathbb{C}^n$ to effectively compute stabilizing controllers of multidimensional systems, as well as the Youla-Kučera parametrization of all the stabilizing controllers [90]. This last parametrization can be used to transform standard H_∞-optimal control problems, which are nonlinear by nature, into affine, and thus convex optimal problems. See [37], [90] and the references therein. Applications addressed in the former ANR MSDOS (2014–2018) will be developed in collaboration with Bachelier (U. Poitiers). The algorithms obtained in this direction will be unified in a unique `Maple` package.

Finally, the *noncommutative geometric approach* to robust problems for infinite-dimensional linear systems (e.g. differential time-delay or PD systems) [54], initiated in [92], will be further studied based on the mathematical concepts and methods introduced by (the Fields medalist) Connes [53]. We particularly want to investigate generalizations of Nyquist's theorem to infinite-dimensional systems based on index theory (pairing of K-theory and K-homology), model reduction based on Connes' interpretation of infinitesimal operators, robustness metrics particularly the ν-gap metric, etc. The quantized differential calculus [53], based on Hankel operators, as well as the connections and curvatures on stabilizable systems will be further studied [92]. We aim to exploit these noncommutative differential geometric structures on the systems to get new inside in both the topology and geometry aspects of the H_∞-control theory for infinite-dimensional systems [54].

3.3.2. Parameter estimation for linear & nonlinear functional systems

3.3.2.1. Linear functional systems

Our expertise on algebraic parameter estimation problem, coming from the former NON-A project-team, will be further developed. Following [65], this problem consists in estimating a set θ of parameters of a signal $x(\theta, t)-$ which satisfies a certain dynamics $-$ when the signal $y(t) = x(\theta, t) + \gamma(t) + \varpi(t)$ is observed, where γ denotes a structured perturbation and ϖ a noise. For instance, x can be a multi-sinusoidal waveform signal and θ phases, frequencies, or amplitudes [13]. Based on a combination of algebraic analysis techniques (rings of differential operators), differential elimination theory (computation of annihilators), and *operational calculus* (Laplace transform, convolution), [65] shows how θ can sometimes be explicitly determined by means of closed-form expressions using iterated integrals of y. These integrals usually help to filter the effect of the noise ϖ on the estimation of the parameters θ.

A first aim in this direction is to develop to a greater extent our recent work [108] that shows how the above approach can cover wider classes of signals such as *holonomic signals* (e.g. signals decomposed into orthogonal polynomial bases, special functions, possibly wavelets).

Moreover, [94] explains how larger classes of structured perturbations γ can be considered when the approach developed in [65], [108], based on computation of *annihilators*, is replaced by a new approach based on the more general algebraic concept of *syzygies* [103]. This general approach to the algebraic parameter estimation problem will be developed. Following the ideas of [94], an effective version of this general approach will also be done based on differential elimination techniques, i.e. Gröbner basis techniques for rings of differential operators. It will be implemented in a dedicated `Maple` package which will extend the current prototype `NonA` package [94].

Furthermore, as an alternative to passing forth and backwards from the time domain to the operational (Laplace/frequency) domain by means of Laplace transform and its inverse as done in the standard algebraic parameter estimation method [65], [108], we aim to develop a direct time domain approach based on calculus on rings of integro-differential operators as described by the following picture (L denotes the Laplace transform):

$$\begin{array}{ccc} \text{temporal domain} & & \text{frequency domain } L(z)=\widehat{z} \\ z(t)=x(t,\theta)+\gamma(t) & \Longrightarrow & \widehat{z}(s)=\widehat{x}(s,\theta)+\widehat{\gamma}(s) \\ \mathbf{integro-diff.\ calculus}\ \Downarrow & & \Downarrow \\ \text{closed-form expressions} & & \text{differential algebraic calculus} \\ \theta = g\left(\int^i z(t)\right) & \Longleftarrow & \theta = f(s^{-i}\,\widehat{z}(s)) \end{array}$$

The direct computation will be handled by means of the effective methods of rings of integro-differential operators described in the above sections.

3.3.2.2. Nonlinear functional systems

For nonlinear control systems, the approach to the parameter estimation problem, recently proposed in [3] and based on the computation of integro-differential input-output equations, will be further developed based on the integration of fractions [2]. Such a representation better suits a numerical estimation of the parameters as shown in [3].

In [29], we have recently initiated an extension of the results developed in [3] to handle integro-differential equations such as Volterra-Kostitzin's equation. This general approach, based on an extension of the input-output ideal method for ordinary differential equations to the integro-differential ones, will be further developed based on the effective elimination theory for systems of integro-differential equations. An important advantage of this approach is that not only it solves the identifiability theoretical question but it also prepares a further parameter estimation step [57].

4. Application Domains

4.1. Adaptive & parametric robust control – collaboration with Safran Electronics & Defense

We have developed a collaboration with *Safran Electronics & Defense* (Massy Palaiseau) and Rouillier (OURAGAN, Inria Paris) on a *parametric robust control theory* based on computer algebra methods (symbolic-numeric methods), as well as its applications to the robust stabilization of certain mechanical systems (e.g. gyrostabilized systems, two mass-spring-damper system, stabilized mirrors).

For low-dimensional systems of ODEs, this approach aims to determine closed-form solutions for robust controllers and for the robustness margins in terms of the model parameters (e.g. mass, length, inertia, mode) [12], [98], [100]. The main applications of these results are twofold: the feasibility of an industrial project can be simplified by speeding up the computation of robust controllers and robust margins for systems with rapidly changing architecture parameters, and avoiding usual time-consuming optimization techniques. Secondly, adaptive and embeddable schemes for robust controllers can be proposed and tested while coupling our approach with real-time parameter estimation methods such as the ones developed in the GAIA team. For more details, see [12].

Preliminary works in the direction have opened a great variety of questions such as the explicit search for positive definite solutions of algebraic or differential Riccati equations (i.e. polynomial or differential systems) with model parameters, the reduction of these equations, and of the parameters based on symmetries, the development, of efficient tools for plotting high degree curves and surfaces showing the robustness margins in terms of the model parameters (collaboration with Moroz (GAMBLE, Inria Nancy)), the use of a certified numeric Newton-Puiseux algorithm for the design of robust controllers, etc. [12], [98], [100]. These results require the use of a large spectrum of computer algebra methods such as linear algebra with parameters, polynomial systems with parameters, ordinary differential systems with parameters, symmetries and reduction, rational parametrizations, discriminant varieties, semi-algebraic sets, critical point methods, real root isolation methods, etc. We shall further develop the parametric robust control in collaboration with *Safran Electronics & Defense*.

In connection with the above results, parameter estimation methods will be studied to develop *adaptive robust controllers* for gyrostabilized systems. Indeed, combining explicit characterizations of robust controllers in terms of the model parameters with time-to-time estimations of these model parameters (which can change with the system production, the heat, the wear, etc.), the robust controllers can then be automatically tuned to conserve their robustness performances [12], [99].

Finally, as explained in [11], [99], constant and distributed delays naturally appear in *Safran E & D* systems (e.g. gyrostabilized systems using visual trackers, stabilized mirror models). Extensions of the above problems and results will be studied for differential time-delay systems based on robust control techniques for infinite-dimensional systems (see, e.g., [54] and the references therein) and its algebraic extension to include model parameters.

4.2. Self calibration problem & Gear fault diagnosis − collaboration with Safran Tech

4.2.1. Self calibration problem

Due to numerous applications (e.g. sensor network, mobile robots), sources and sensors localization has intensively been studied in the literature of signal processing. The *anchor position self calibration problem*, a well-known problem in signal processing, consists in estimating the positions of both the moving sources and a set of fixed sensors (anchors) when only the distance information between the points from the two different sets is available. The position self-calibration problem is a particular case of the *Multidimensional Unfolding* (MDU) problem for the Euclidean space of dimension 3.

Based on computer algebra methods for polynomial systems, we have recently proposed a new approach for the MDU problem which yields closed-form solutions and an efficient algorithm for the estimation of the positions [56] only based on linear algebra techniques. This first result, obtained in collaboration with Dagher (Research Engineer, Inria Chile) and Zheng (DEFROST, Inria Lille - Nord Europe), yields a recent *patent* [55]. Real tests are now carried out. Our first results will be further developed, improved, tested, and demonstrated.

The MDU problem is just one instance of localization problems: more problems can be addressed for which a computer algebra expertise can brought new interesting results, especially in finding closed-form solutions, yielding new estimation techniques which avoid the use of optimization algorithms as commonly done in the signal processing literature. The main differences between these localization problems can essentially be read on a certain matrix of distance called the *Euclidean distance matrix* [56].

4.2.2. Gear fault diagnosis

We have a collaboration with Barau (*Safran Tech*) and Hubert (*Safran Tech*), and Dagher (Research Engineer, Inria Chile) on the symbolic-numeric study of the new multi-carrier demodulation method developed in [71]. *Gear fault diagnosis* is an important issue in aeronautics industry since a damage in a gearbox, which is not detected in time, can have dramatic effects on the safety of a plane.

Since the vibrations of a spur gear can be modeled as a product of two periodic functions related to the gearbox kinematic, [71] has proposed to recover each function from the global signal by means of an optimal reconstruction problem which, by means of Fourier analysis, can be rewritten as

$$\mathrm{argmin}_{u \in \mathbb{C}^n, v_1, v_2 \in \mathbb{C}^m} \|M - u\, v_1^{\star} - D\, u\, v_2^{\star}\|_F,$$

where $M \in \mathbb{C}^{n \times m}$ (resp. $D \in \mathbb{C}^{n \times n}$) is a given (resp. diagonal) matrix with a special shape, $\| \cdot \|_F$ denotes the Frobenius norm, and $v^{\star}$ the Hermitian transpose of v. Based on closed-form solutions of the exact problem − which are defined by a system of polynomial equations in the unknowns − we have recently proposed efficient numerical algorithms to numerically solve the problem. The first results are interesting and they will be further developed and tested on different data sets. Finally, we shall continue to study the extremal solutions of the corresponding polynomial problem by means of symbolic and numeric methods, etc.

4.3. Applications of the parameter estimation problem to multidisciplinary domains − collaboration with an INSERM team (Rouen University)

For linear systems, the closed-form expressions of the parameters obtained by means of the algebraic parameter estimation problem will continue to provide robust estimates in our multidisciplinary collaborations, in marine biology and human-machine interactions, as it is already the case of existing NON-A results (in collaboration with LOKI, Inria Lille - Nord Europe).

For nonlinear systems, a collaboration with biologists and modelers has been developed for a few years already [29]. Our partners are a team from the Applied Mathematical Department of Le Havre University (modelers) and an INSERM team at Rouen University (neurobiologists). The targeted biological problem is the cortical spreading depression, a brain disease likely to occur after cerebrovascular accidents [60]. We seek − ultimately − a mathematical integro-differential model permitting to predict the triggering of this disease for patients arising in the emergency services of hospitals. The key phenomenon to reproduce is a slow depolarization wave of neurons. Our approach is original because it focuses on the role calcium fluxes in neurons and astrocytes.

5. Highlights of the Year

5.1. Highlights of the Year

Computer Algebra in Scientific Computing

The GAIA team organized the conference *Computer Algebra in Scientific Computing* (CASC), University of Lille, 17–21 September 2018.

6. New Software and Platforms

6.1. ADHOMFI

Adaptive Homogeneous Filtering
KEYWORDS: Automatic differentiation - Filtering

FUNCTIONAL DESCRIPTION: allows to reconstruct a signal based on derivatives estimation and to filter high amplitude and wide frequencies spectrum perturbations.

- Contact: Denis Efimov

6.2. Platforms

6.2.1. BLINEIDE library

We have released the first version of the BLINEIDE library. BLINEIDE stands for Bibliothèques Lilloises dédiées à l'Intégration Numérique des Équations Intégro-Différentielles. It is an open source C library dedicated to the numerical integration of systems of integro-differential equations. for more details, see https://pro.univ-lille.fr/francois-boulier/logiciels/blineide.

7. New Results

7.1. Regular (differential) chains

[17] provides new equivalence theorems for regular chains and regular differential chains, which are generalizations of Ritt's characteristic sets. These theorems focus on regularity properties of elements of residue class rings defined by these chains, which are revealed by resultant computations. New corollaries to these theorems have quite simple formulations.

[30] contains a description of the management of the parameters in the`Maple DifferentialAlgebra package` and, in particular, in the RosenfeldGroebner function.

7.2. Systems of integro-differential equations

[28], [29] present a proof of concept for symbolic and numeric methods dedicated to the parameter estimation problem for models formulated by means of nonlinear integro-differential equations. In particular, we address the computation of the model input-output equation and the numerical integration of integro-differential systems (the BLINEIDE library).

7.3. Certified non-conservative tests for the structural stability of discrete multidimensional systems

In collaboration with Fabrice Rouillier (Inria Paris, Ouragan), in [18], we propose a new approach for testing the stability of nD systems. We first 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 the fact that a certain system of polynomials does not have real solutions. We then use state-of-the-art computer algebra algorithms to check this last condition, and thus the structural stability of multidimensional systems. Our results have been implemented in a `Maple` prototype.

7.4. Using symbolic computation to solve algebraic Riccati equations arising in invariant filtering

In this joint work with Axel Barrau from Safran Tech [23], we propose a new step in the development of invariant observers. In the past, this theory led to impressive simplifications of the error equations encountered in estimation problems, especially those related to navigation. This was used to reduce computation load or derive new theoretical properties. Here, we leverage this advantage to obtain closed-form solutions of the underlying algebraic Riccati equations through advanced symbolic computation methods.

7.5. Parametric sub-optimal H_∞ controllers for an optro-mechanical system

In collaboration with *Safran Electronics & Defense*, in [15], we studied the robust stabilization of the line of sight of a stabilized mirror system. This system can be modeled by a single-input single- output time-delay system. Due to large model uncertainties, non-parametric methods are usually too conservative. Hence, we consider here unfixed model parameters. Using an additive decomposition, we show how to compute parametric H_∞ controllers of the time-delay model. Such a symbolic approach is interesting in the context of adaptive control and is illustrated throughout a simulation with an ideal parameter estimator.

7.6. A symbolic approach for signal demodulation and application to gearbox vibration analysis

This work is made in collaboration with Axel Barrau and Elisa Hubert (Safran Tech), and Roudy Dagher (Research Engineer, Inria Chile). The problem under study, which reduces to a certain signal factorization problem, was shown by Barrau et. al. to be equivalent to a Frobenius norm minimization problem. Starting from this optimization problem, we investigate the use of computer algebra methods to compute explicit solutions for the original problem. Along the way, we exhibit interesting algebraic and geometric properties of the underlying polynomial system. A paper is currently in development to summarize these results.

7.7. Curve analysis for the stability of time-delay systems

This work aims to design a new symbolic-numerical Puiseux-free approach for the study of the stability of differential time-delay systems. The idea behind is to replace the costly computations of Puiseux developpements around the *critical pairs* of the characteristic function by the numerical analysis of the branches of a well chosen 3D curve. The preliminary results show that this approach is easier to implement and turns out to be more efficient in practice. This ongoing work will be the subject of a future publication.

8. Bilateral Contracts and Grants with Industry

8.1. Safran Electronics & Defense

Within the CIFRE PhD thesis (2014-2018) [15], we have studied new robust stabilization techniques for gyrostabilized systems with unfixed model parameters (e.g. modes, masses, stiffness of springs, damper magnitudes). Parameters of their models indeed slowly change with the temperature, fatigue, etc., yielding time-consuming re-computations of robust controllers. Moreover, the possibility to quickly know robustness indicators (e.g. margins) and explicit robust controllers in terms of the model parameters can highly speed up the design of a project. Finally, closed-form solutions for robust controllers in terms of the model parameters are the first steps towards the development of adaptive robust controllers which can be embedded in gyrostabilized platforms since no optimization algorithms are then required for a real-time implementation and only the parameters have to be estimated from time to time to re-compute the robust controller (based on a basic arithmetic). To do that, we have introduced algebraic methods and computer algebra techniques to initiate a new approach entitled *parametric robust control*. For mor details, see [15] and [98], [100], [99]. This new approach will be further developed in the future since it opens both theoretical and practical interesting questions. In particular, the new PhD thesis of Grace in GAIA aims to study the underlying mathematical problem from both a theoretical and an implementation perspectives.

8.2. Ellcie Healthy

A new collaboration with Ellcie Healthy, a company based in Nice began in October 2017. It involves the analyze of signals coming from optical sensors installed in glasses. With Denis Efimov, the first studies obtained were very promising. This collaboration was formalized with the signature of a first contract in March 2018. The first objective of this project was to design algorithms for intelligent filtering of data coming from infrared sensors, especially for light-related disturbances. Discussions are currently underway for the submission of new joint projects.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

- *ANR project MSDOS* (Multidimensional System: Digression on Stability, coordinator: Nima Yeganefar (Poitiers University), 2014-2018) aimed at studying stability and stabilization problems for multidimensional systems by means of both analytic and algebraic methods. For more information, see https://www.lias-lab.fr/msdos/doku.php.
- *ANR TurboTouch* (High-performance touch interactions, coordinator: G. Casiez (MJOLNIR team, Inria), 2014–2019) develops methods and tools on transfer functions to allow high performance tactile interactions (e.g. high precision and low latency) adapted to the user and to the task. This research project is developed in collaboration with the Loki team, Inria Lille – Nord Europe (project leader). For more information, seehttp://mjolnir.lille.inria.fr/turbotouch/.
- *ANR WaQMoS* (Coastal waters Quality surveillance using bivalve Mollusk-based Sensors, coordinator: D. Efimov (Non-A Post, Inria), 2015–2020) develops a biosensor, based on measurements and interpretation of bi- valves mollusks behavior, for remote online detection of coastal water pollution and cli- mate change consequences. This research project is developed in collaboration with the Valse team, Inria Lille – Nord Europe (project leader). For more information, see https://team.inria.fr/non-a/anr-waqmos/.

9.2. European Initiatives

9.2.1. Collaborations with Major European Organizations

Mohamed Barakat: University of Siegen (Germany)

Effective module theory, effective homological algebra, algebraic analysis, computer algebra, implementation.

Georg Regensburger: Institute for Algebra, Johannes Kepler University Linz (Austria)

Rings of integro-differential-delay operators, computer algebra, implementation.

Daniel Robertz: University of Plymouth (United Kingdom)

Effective algebraic analysis, mathematical systems theory, computer algebra, implementation.

9.3. International Initiatives

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

WeCare, Inria Northern European Associate Team with the team of A. Medvedev from Uppsala University on effective algorithms for estimation and control in wearable devices for health and care, 2018–2020.

We participate in *HoTSMoCE*, an Inria Associated team with Non-A Post and the team of L. Fridman (UNAM, Mexico), on the development of algebraic and homogeneous tools for sliding mode control and estimation.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Thomas Cluzeau, XLIM, University of Limoges, May 2018.
- Marc Moreno Maza, University of Western Ontario, London, Ontario, Canada, September 2018.
- Alexander Medvedev, University of Uppsala (03–05/10/2018).
- Fredrik Olsson, University of Uppsala (26–30/11/2018).
- Elisa Hubert (Safran Tech) visited us twice (23–24/07/2018, 12–13/09/2018) to work on the problem of gear fault diagnostic based on algebraic and symbolic approaches.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

The GAIA team organized the 2018 edition of the *Computer Algebra in Scientific Computing* (CASC) international workshops series (17–21 September 2018). For more details, see http://www.casc.cs.uni-bonn.de/2018/.

10.1.1.1. General Chair, Scientific Chair

- Since 2018, François Boulier is a *General Chair* of CASC. See http://www.casc-conference.org/.
- A. Quadrat is a member of the *IFAC Technical Committee* “Linear Control Systems”, International Federation of Automatic Control, TC2.2.

10.1.1.2. Member of the Organizing Committees

The GAIA team organized the *Computer Algebra in Scientific Computing* (CASC), University of Lille, 17–21 September 2018.

A. Quadrat is a member of the organization committee of the *Journées Nationales de Calcul Formel* (JNCF), Luminy, France, 22–26/01/2018.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

F. Lemaire was Poster Chair of the *International Symposium on Symbolic and Algebraic Computation* (ISSAC'2018) Poster Session, New York, 16–19/05/2018.

10.1.2.2. Reviewer

- Y. Bouzidi reviewed a publication for ISSAC 2018 and for TDS 2018.
- F. Lemaire reviewed a publication for CASC 2018.
- A. Quadrat reviewed two publications for ISSAC 2018.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

A. Quadrat is associate editor of *Multidimensional Systems and Signal Processing*.

10.1.3.2. Reviewer - Reviewing Activities

A. Quadrat reviewed papers for the journal *Multidimensional Systems and Signal Processing*.

10.1.4. Invited Talks

Y. Bouzidi gave the following talks:

- A Symbolic Approach for Solving Algebraic Riccati Equations, *Journées Nationales de Calcul Formel* (JNCF), CIRM, Marseille - January 22th 2018.
- Parametric study of the critical pairs of linear differential systems with commensurate delays, *Inria DISCO team-project*, Paris - April 30th 2018.
- Using symbolic computation to solve algebraic Riccati equations arising in invariant filtering, *European Control Conference*, Limassol - June 15th 2018.
- A symbolic approach for a parametric H_∞ control problem, *visualization group*, ISCD, Paris - Novembre 30th 2018.

10.1.5. Leadership within the Scientific Community

A. Quadrat co-organized with N. Yeganefar (University of Poitiers) the invited session *New Results in Multidimensional Systems Theory* at the *2018 European Control Conference*, Cyprus, June 12-15, 2018.

R. Ushirobira was co-responsible with Denis Efimov (Non-A Post, Inria) and Gilberto Pin (Electrolux Italia) of a special issue Finite-time estimation, diagnosis and synchronization of uncertain systems for the *European Journal of Control*.

10.1.6. Scientific Expertise

A. Quadrat is a member of the *Bureau du Comité des Equipes-Projets* (BCEP) and of the *Commission des Emplois de Recherche*, Inria Lille - Nord Europe.

A. Quadrat wrote a comparative report for the position "Algorithmische Algebra" (W2), University of Siegen, Germany.

10.1.7. Research Administration

- F. Boulier has been Head of the *Spécialité GIS at Polytech Lille* since June 2018.
- A. Quadrat is in charge with C. Jamroz of the *RaWeb 2018* for Inria Lille – Nord Europe.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Rosane Ushirobira taught around 90h at Polytech Lille and University of Lille (Linear algebra, analysis and logic).

10.2.2. Supervision

PhD: Guillaume Rance, "Commande H_∞ paramétrique et application aux viseurs gyrostabilisés", University Paris - Saclay, 09/07/2018, Alban Quadrat & Arnaud Quadrat & Hugues Mounier.

PhD in progress: Grace Younes, "Calcul de multiplicités de racines de polynoômes et de quasi-polynoômes", 01 /10/2018, Alban Quadrat & Yacine Bouzidi & Adrien Poteaux.

Master: Ambroise Fleury (Licence 3 ème année training), "Computation of gcd using AVX", F. Lemaire.

Master: Guillaume Maitrot (Master 2 training), "Improving the BLINEIDE library with OpenMP", F. Lemaire.

10.2.3. Juries

F. Boulier was referee of the PhD thesis of G. Rance, University Paris Orsay, 09/07/2018.

A. Quadrat was a jury member of the CRCN Inria 2018 competition for the Lille – Nord Europe center.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Since December 2103, R. Ushirobira organize the cycle “30 minutes of science”, a rotating monthly seminar for all researchers at Inria LIlle. On average, 40 people participate in this seminar.

10.3.2. Education

In 2017/2018, Rosane Ushirobira was a referent researcher for *the Math en Jean program* at Arthur Rimbaud College (Villeneuve d’Ascq).

10.3.3. Internal action

Y. Bouzidi, *Symbolic-numeric method for a parametric control problem*, presentation in the *30 minutes of science Inria event*, Lille - January 10th 2018.

11. Bibliography

Major publications by the team in recent years

[1] F. BOULIER, C. CHEN, F. LEMAIRE, M. MORENO-MAZA.*Real root isolation of regular chains*, in "Computer Mathematics", 2014, p. 33–38

[2] F. BOULIER, F. LEMAIRE, J. LALLEMAND, G. REGENSBURGER, M. ROSENKRANZ.*Additive normal forms and integration of differential fractions*, in "Journal of Symbolic Computation", 2016, vol. 77, p. 16–38

[3] F. BOULIER, F. LEMAIRE, M. ROSENKRANZ, R. USHIROBIRA, N. VERDIÈRE.*On symbolic approaches to integro-differential equations*, Springer, 2019, p. 165–184

[4] Y. BOUZIDI, S. LAZARD, G. MOROZ, M. POUGET, F. ROUILLIER.*Separating linear forms and rational univariate representations of bivariate systems*, in "Journal of Symbolic Computation", 2015, vol. 68, p. 84–119

[5] Y. BOUZIDI, A. POTEAUX, A. QUADRAT.*A symbolic computation approach towards the asymptotic stability analysis of differential systems with commensurate delays*, in "Delays and Interconnections: Methodology, Algorithms and Applications, Advances in Delays and Dynamics (ADD)", Springer, 2018

[6] Y. BOUZIDI, A. QUADRAT, F. ROUILLIER.*Certified Non-conservative Tests for the Structural Stability of Discrete Multidimensional Systems*, in "Multidimensional Systems and Signal Processing", June 2018, https://hal.inria.fr/hal-01951765

[7] F. CHYZAK, A. QUADRAT, D. ROBERTZ.*Effective algorithms for parametrizing linear control systems over Ore algebras*, in "Applicable Algebra in Engineering, Communications and Computing", 2005, vol. 16, p. 319–376

[8] T. CLUZEAU, A. QUADRAT.*Factoring and decomposing a class of linear functional systems*, in "Linear Algebra and Its Applications", 2008, vol. 428, p. 324–381

[9] A. POTEAUX, M. RYBOWICZ.*Good reduction of Puiseux series and applications*, in "Journal of Symbolic Computation", 2012, vol. 47, p. 32–63

[10] A. POTEAUX, M. RYBOWICZ.*Improving complexity bounds for the computation of Puiseux series over finite fields*, in "ISSAC", 2015, p. 299–306

[11] A. QUADRAT, A. QUADRAT.*Delay effects in visual tracking problems for an optronic sighting system*, in "Low-Complexity Controllers for Time-Delay Systems, Advances in Delays and Dynamics (ADD)", Springer, 2014, vol. 2, p. 77–92

[12] G. RANCE, Y. BOUZIDI, A. QUADRAT, A. QUADRAT.*H_∞ parametric control via algebraic Riccati equations*, in "22nd International Symposium on Mathematical Theory of Networks and Systems (MTNS", 2016

[13] 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, vol. 30, p. 1689–1713

[14] R. USHIROBIRA, A. QUADRAT.*Algebraic estimation of a biased and noisy continuous signal via orthogonal polynomials*, in "55th IEEE Conference on Decision and Control (CDC)", 2016

Publications of the year

Doctoral Dissertations and Habilitation Theses

[15] G. RANCE.*Parametric H∞ control and its application to gyrostabilized sights*, Université Paris-Saclay, July 2018, https://tel.archives-ouvertes.fr/tel-01904086

Articles in International Peer-Reviewed Journal

[16] H. AHMED, R. USHIROBIRA, D. EFIMOV.*Robust global synchronization of Brockett oscillators*, in "IEEE Transactions on Control of Network Systems", 2018, p. 1-10 [*DOI :* 10.1109/TCNS.2018.2813927], https://hal.inria.fr/hal-01965539

[17] F. BOULIER, F. LEMAIRE, A. POTEAUX, M. MORENO MAZA.*An Equivalence Theorem For Regular Differential Chains*, in "Journal of Symbolic Computation", 2018, p. 1-27, https://hal.archives-ouvertes.fr/hal-01391768

[18] Y. M. BOUZIDI, A. QUADRAT, F. ROUILLIER.*Certified Non-conservative Tests for the Structural Stability of Discrete Multidimensional Systems*, in "Multidimensional Systems and Signal Processing", June 2018, https://hal.inria.fr/hal-01951765

[19] X. DERU, F. LEMAIRE, D. NICOLAS.*From Etaples to Pompei. Discussion about Metal Vessel in Roman Household*, in "Herom", 2018, Définition d'un modèle à partir des données écrites et archéologiques du camp napoléonien d'Etaples, confrontation des données céramologiques romaines en Gaule et en Italie, https://hal.univ-lille3.fr/hal-01737593

[20] D. EFIMOV, R. USHIROBIRA, J. A. MORENO, W. PERRUQUETTI.*Homogeneous Lyapunov functions: from converse design to numerical implementation*, in "SIAM Journal on Control and Optimization", December 2018, https://hal.inria.fr/hal-01851761

[21] J. G. RUEDA-ESCOBEDO, R. USHIROBIRA, D. EFIMOV, J. MORENO.*Gramian-based uniform convergent observer for stable LTV systems with delayed measurements*, in "International Journal of Control", 2018, https://hal.inria.fr/hal-01889193

[22] H. SILM, R. USHIROBIRA, D. EFIMOV, J.-P. RICHARD, W. MICHIELS.*A note on distributed finite-time observers*, in "IEEE Transactions on Automatic Control", May 2018 [*DOI :* 10.1109/TAC.2018.2838042], https://hal.inria.fr/hal-01827041

International Conferences with Proceedings

[23] A. BARRAU, G. RANCE, Y. BOUZIDI, A. QUADRAT, A. QUADRAT.*Using symbolic computation to solve algebraic Riccati equations arising in invariant filtering*, in "ECC2018 - European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01667341

[24] P.-A. BLIMAN, D. EFIMOV, R. USHIROBIRA.*A class of nonlinear adaptive observers for SIR epidemic model*, in "ECC 2018 - European Control Conference", Limassol, Cyprus, June 2018, 6, https://hal.inria.fr/hal-01724989

[25] M. NANCEL, S. ARANOVSKIY, R. USHIROBIRA, D. EFIMOV, S. POULMANE, N. ROUSSEL, G. CASIEZ.*Next-Point Prediction for Direct Touch Using Finite-Time Derivative Estimation*, in "Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2018)", Berlin, Germany, October 2018 [*DOI :* 10.1145/3242587.3242646], https://hal.inria.fr/hal-01893310

[26] J. G. RUEDA-ESCOBEDO, R. USHIROBIRA, D. EFIMOV, J. MORENO.*A Gramian-based observer with uniform convergence rate for delayed measurements*, in "ECC 2018 - 16th annual European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01895389

[27] H. SILM, R. USHIROBIRA, D. EFIMOV, W. MICHIELS, J.-P. RICHARD, E. FRIDMAN.*Comparison of the Time-Delay Margin of a Distributed and Centralized Observer*, in "ECC 2018 - European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01761390

Conferences without Proceedings

[28] F. BOULIER.*On the Parameter Estimation Problem for Integro-Differential Models*, in "AADIOS @ ACA 2018", Santiago de Compostela, Spain, June 2018, https://hal.archives-ouvertes.fr/hal-01780909

[29] F. BOULIER, H. CASTEL, N. CORSON, V. LANZA, F. LEMAIRE, A. POTEAUX, A. QUADRAT, N. VERDIÈRE.*Symbolic-Numeric Methods for Nonlinear Integro-Differential Modeling*, in "CASC 2018 - The 20th International Workshop on Computer Algebra in Scientific Computing", Lille, France, September 2018, https://hal.archives-ouvertes.fr/hal-01765409

Other Publications

[30] F. BOULIER.*The Management of Parameters in the MAPLE DifferentialAlgebra Package*, June 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01825191

[31] F. BOULIER, F. FAGES, O. RADULESCU, S. S. SAMAL, A. SCHUPPERT, W. M. SEILER, T. STURM, S. WALCHER, A. WEBER.*The SYMBIONT Project: Symbolic Methods for Biological Networks*, August 2018, F1000Research 7:1341 (poster), https://hal.inria.fr/hal-01889825

[32] G. PEROZZI.*Optimization of the control effort for quadrotors using a varying-gain quasi-continuous sliding mode control influenced by a wind estimator*, April 2018, working paper or preprint, https://hal.inria.fr/hal-01758112

References in notes

[33] S. S. ABHYANKAR.*Irreducibility criterion for germs of analytic functions of two complex variable*, in "Advances in Mathematics", 1989, vol. 74, p. 190–257

[34] J. ACZÉL.*Lectures on Functional Equations and Their Applications*, Academic Press, 1966

[35] J. ACZÉL, J. DHOMBRES.*Functional Equations in Several Variables*, Cambridge University Press, 1989

[36] M. AUDIN.*Hamiltonian Systems and Their Integrability*, American Mathematical Society, 2008

[37] J. A. BALL, S. TER HORST.*Robust control, multidimensional systems and multivariable Nevanlinna-Pick interpolation*, Topics in operator theory. Volume 2. Systems and Mathematical Physics, Birkhäuser, 2010, p. 13–88

[38] J.-D. BAUCH, E. NART, H. STAINSBY.*Complexity of the OM factorizations of polynomials over local fields*, in "LMS Journal of Computation and Mathematics", 2013, vol. 16, p. 139–171

[39] V. V. BAVULA.*The algebra of integro-differential operators on an affine line and its modules*, in "J. Pure Appl. Algebra", 2013, vol. 217, p. 495–529

[40] T. BECKER, V. WEISPFENNING.*Gröbner Bases. A Computational Approach to Commutative Algebra*, Springer, 1993

[41] J. E. BJÖRK.*Rings of Differential Operators*, North Holland, 1979

[42] N. K. BOSE.*Multidimensional Systems Theory - Progress, Directions and Open Problems in Multidimensional Systems*, Reidel Publishing Company, 1985

[43] N. K. BOSE.*Multidimensional Systems Theory and Applications*, Springer, 1995

[44] F. BOULIER.*BLAD: Bibliothèques Lilloises d'Algèbre Différentielle*, http://cristal.univ-lille.fr/~boulier/pmwiki/pmwiki.php/Main/BLAD

[45] F. BOULIER.*An optimization of Seidenberg's elimination algorithm in differential algebra*, in "Mathematics and Computers in Simulation", 1996, vol. 42, p. 439–448

[46] F. BOULIER, D. LAZARD, F. OLLIVIER, M. PETITOT.*Computing representations for radicals of finitely generated differential ideals*, in "Appl. Algebra Engrg. Comm. Comput.", 2009, vol. 20, p. 73–121

[47] Y. BOUZIDI, T. CLUZEAU, G. MOROZ, A. QUADRAT.*Computing effectively stabilizing controllers for a class of nD systems*, in "Proc. of the 20th World Congress of the International Federation of Automatic Control (IFAC'17), 9-14 July 2017", 2017

[48] Y. BOUZIDI, A. QUADRAT, F. ROUILLIER.*Computer algebra methods for testing the structural stability of multidimensional systems*, in "IEEE 9th International Workshop on Multidimensional Systems (NDS)", 2015

[49] D. BRIDGES, R. MINES, F. RICHMAN, P. SCHUSTER.*The polydisk Nullstellensatz*, in "Proceedings of the American Mathematical Society", 2003, vol. 132, p. 2133–2140

[50] B. BUCHBERGER.*Gröbner bases: An algorithmic method in polynomial ideal theory*, in "Multidimensional Systems Theory: Progress, Directions and Open Problems in Multidimensional Systems", N. K. BOSE (editor), Reidel, 1985, p. 184–232

[51] E. CARTAN.*Les systèmes différentiels extérieurs et leurs applications géométriques*, Hermann, 1971

[52] X. CARUSO, D. ROE, T. VACCON.*Division and slope factorization of p-adic polynomials*, in "International Symposium on Symbolic and Algebraic Computation (ISSAC)", 2016

[53] A. CONNES.*Noncommutative Geometry*, Academic Press, 1994

[54] R. F. CURTAIN, H. J. ZWART.*An Introduction to Infinite-Dimensional Linear Systems Theory*, Springer, 1995

[55] R. DAGHER, A. QUADRAT, G. ZHENG.*Auto-localisation par mesure de distances*, Pattern FR1853553, Inria France, 2018

[56] R. DAGHER, A. QUADRAT, G. ZHENG.*Algebraic solutions to the metric multidimensional unfolding. Application to the position self-calibration problem*, in "Pattern under development", 2017

[57] L. DENIS–VIDAL, G. JOLY–BLANCHARD, C. NOIRET.*System identifiability (symbolic computation) and parameter estimation*, in "Numerical Algorithms", 2003, vol. 34, p. 282—292

[58] S. DIOP.*Elimination in control theory*, in "Math. Control Signals Systems", 1991, vol. 4, p. 17–32

[59] S. DIOP.*Differential-algebraic decision methods and some applications to system theory*, in "Theoret. Comput. Sci.", 1992, vol. 98, p. 137–161

[60] J. P. DREIER, S. MAJOR, A. MANNING, J. WOITZIK, C. DRENCKHAHN, J. STEINBRINK, C. TOLIAS, A. I. OLIVEIRA-FERREIRA, M. FABRICIUS, J. A. HARTINGS, P. VAJKOCZY, M. LAURITZEN, U. DIRNAGL, G. BOHNER, A. J. STRONG.*Cortical spreading ischaemia is a novel process involved in ischaemic damage in patients with aneurysmal subarachnoid haemorrhage*, in "Brain", 2009, vol. 132, p. 1866–1881

[61] D. DUVAL.*Rational Puiseux expansions*, in "Compositio Mathematica", 1989, vol. 70, p. 119–154

[62] A. FABIAŃSKA, A. QUADRAT.*Applications of the Quillen-Suslin theorem to multidimensional systems theory*, in "Gröbner Bases in Control Theory and Signal Processing, Radon Series on Computation and Applied Mathematics", D. GRUYTER (editor), H. Park, G. Regensburger, 2007, vol. 3, p. 23–106, http://chercheurs.lille.inria.fr/aquadrat/QuillenSuslin.html

[63] M. FLIESS, S. T. GLAD.*An algebraic approach to linear and nonlinear control*, in "Essays on Control: Perspectives in the Theory and its Applications", BIRKHAÜSER (editor), 1993, p. 223–267

[64] M. FLIESS, H. MOUNIER.*Controllability and observability of linear delay systems: an algebraic approach*, in "ESAIM Control Optim. Calc. Var.", 1998, vol. 3, p. 301–314

[65] M. FLIESS, H. SIRA-RAMÍREZ.*An algebraic framework for linear identification*, in "ESAIM Control Optim. Calc. Variat.", 2003, vol. 9, p. 151–168

[66] H. GLUESING-LUERSSEN.*Linear Delay-Differential Systems with Commensurate Delays: An Algebraic Approach*, Lecture Notes in Mathematics, Springer, 1995, vol. 1770

[67] M. GROMOV.*Partial Differential Relations*, Springer, 1986

[68] J. GUÀRDIA, J. MONTES, E. NART.*Newton polygons of higher order in algebraic number theory*, in "Transactions of the American Mathematical Society", 2012, vol. 364, p. 361–416

[69] E. HAIRER, G. WANNER.*Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Springer Series in Computational Mathematics, Springer, 2010, vol. 14

[70] J. K. HALE, S. M. V. LUNEL.*Introduction to Functional Differential Equations*, Applied Mathematical Sciences, Springer, 1993, vol. 99

[71] E. HUBERT, A. BARRAU, M. E. BADAOUI.*New multi-carrier demodulation method applied to gearbox vibration analysis*, in "2018 IEEE International Conference on Acoustics, Speech, and Signal, Calgary, 15-20/04/2018", 2018

[72] E. HUBERT.*Notes on triangular sets and triangulation-decomposition algorithms II: Differential Systems*, in "Symbolic and Numerical Scientific Computations, LNCS", SPRINGER (editor), Springer, 2003, vol. 2630, p. 40–87

[73] T. KAILATH.*Linear Systems*, Prentice-Hall, 1980

[74] F. KAKO, T. SASAKI.*Solving multivariate algebraic equations by Hensel construction*, in "Japan J. of Industrial and Applied Math.", 1999, vol. 16, p. 257–285

[75] M. KASHIWARA.*Algebraic Study of Systems of Partial Differential Equation*, Mémoires de la Société Mathématiques de France, 1995, vol. 63, Transl. from Japanese of his 1970 Master's Thesis

[76] E. R. KOLCHIN.*Differential Algebra and Algebraic Groups*, Academic Press, 1973

[77] T. Y. LAM.*Serre's Problem on Projective Modules*, Monograph in Mathematics, Springer, 2006

[78] J. J. LOISEAU.*Algebraic tools for the control and stabilization of time-delay systems*, in "IFAC Reviews, Annual Reviews in Control", 2000, vol. 24, p. 135–149

[79] J. C. MCCONNELL, J. C. ROBSON.*Noncommutative Noetherian Rings*, American Mathematical Society, 2000.

[80] U. OBERST.*Multidimensional constant linear systems*, in "Acta Appl. Math.", 1990, vol. 20, p. 1–175

[81] P. OLVER.*Applications of Lie Groups to Differential Equations*, Springer, 1993

[82] A. PARISA, A. MASOUD, M. M. MAZA.*On the extended Hensel construction and its application to the computation of limit points*, in "International Symposium on Symbolic and Algebraic Computation (ISSAC)", 2017

[83] J. W. POLDERMAN, J. C. WILLEMS.*Introduction to Mathematical Systems Theory. A Behavioral Approach*, TAM, Springer, 1998, vol. 26

[84] J.-F. POMMARET.*Partial Differential Control Theory. Volume II: Control Systems*, Kluwer, 2001

[85] J.-F. POMMARET.*Systems of Partial Differential Equations and Lie Pseudogroups*, Gordon and Breach, 1978

[86] J.-F. POMMARET.*Partial Differential Equations and Group Theory: New Perspectives for Applications*, Kluwer, 1994

[87] A. POTEAUX.*Calcul de développements de Puiseux et application au calcul du groupe de monodrmie d'une courbe algébrique plane*, University of Limoges, 2008

[88] A. POTEAUX, M. WEIMANN.*A dichotomic Newton-Puiseux algorithm using dynamic evaluation*, in "Preprint available at https://hal.inria.fr/hal-01578214", 2017

[89] A. QUADRAT.*The fractional representation approach to synthesis problems: an algebraic analysis viewpoint. Part I: (weakly) doubly coprime factorizations, Part II: internal stabilization*, in "SIAM J. Control & Optimization", 2003, vol. 42, p. 266–299, 300–320

[90] A. QUADRAT.*A lattice approach to analysis and synthesis problems. On a generalization of the Youla-Kučera parametrization. Part II: The lattice approach to MIMO systems*, in "Mathematics of Control, Signals, and Systems", 2006, vol. 18, p. 147–186, 199–235

[91] A. QUADRAT.*An introduction to constructive algebraic analysis and its applications*, in "Les cours du CIRM, Journées Nationales de Calcul Formel", CIRM, 2010, vol. 1, p. 281–471

[92] A. QUADRAT.*Noncommutative geometric structures on stabilizable infinite-dimensional linear systems*, in "European Control Conference, Strasbourg (France)", 2014

[93] A. QUADRAT.*A constructive algebraic analysis approach to Artstein's reduction of linear time-delay systems*, in "Proceedings of 12th IFAC Workshop on Time Delay Systems (TDS)", 2015

[94] A. QUADRAT.*Towards an effective study of the algebraic parameter estimation problem*, in "Proceedings of 20th World Congress of the International Federation of Automatic Control (IFAC'17)", 2017

[95] A. QUADRAT, G. REGENSBURGER.*Computing polynomial solutions and annihilators of integro-differential operators with polynomial coefficients*, in "Algebraic and Symbolic Computation Methods in Dynamical Systems", Springer, 2018, vol. 9, p. 87–111

[96] A. QUADRAT, D. ROBERTZ.*Computation of bases of free modules over the Weyl algebras*, in "Journal of Symbolic Computation", 2007, vol. 42, p. 1113–1141

[97] A. QUADRAT, D. ROBERTZ.*A constructive study of the module structure of rings of partial differential operators*, in "Acta Applicandæ Mathematicæ", 2014, vol. 133, p. 187–243

[98] G. RANCE, Y. BOUZIDI, A. QUADRAT, A. QUADRAT.*Explicit H_∞ controllers for 1st to 3rd order single-input single-output systems with parameters*, in "Proceedings of IFAC 2017 Workshop Congress", 2017

[99] G. RANCE, Y. BOUZIDI, A. QUADRAT, A. QUADRAT.*Parametric sub-optimal H_∞ controllers for an optro-mechanical system modeled by a time-delay 4th order system*, in "8th International Symposium on Optronics in Defence and Security, OPTRO2018, accepted", 2018

[100] G. RANCE, Y. BOUZIDI, A. QUADRAT, A. QUADRAT, F. ROUILLIER.*Explicit H_∞ controllers for 4th order single-input single-output systems with parameters and their applications to the two mass-spring system with damping*, in "Proceedings of IFAC 2017 Workshop Congress", 2017

[101] J. F. RITT.*Differential Algebra*, RI: Amer. Math. Soc., 1950

[102] D. ROBERTZ.*Formal Algorithmic Elimination for PDEs*, Lecture Notes in Mathematics, Springer, 2014, vol. 2121

[103] J. J. ROTMAN.*An Introduction to Homological Algebra*, Academic Press, 1979

[104] W. M. SEILER.*Involution. The Formal Theory of Differential Equations and its Applications in Computer Algebra*, Springer, 2010

[105] J. T. STAFFORD.*Module structure of Weyl algebras*, in "J. London Math. Soc.", 1978, vol. 18, p. 429–442

[106] T. SÁNCHEZ, J. A. MORENO.*A constructive Lyapunov function design method for a class of homogeneous systems*, in "53rd IEEE Conference on Decision and Control, Los Angeles", 2014, p. 5500–5505

[107] T. SÁNCHEZ, J. A. MORENO.*An SOS method for the design of continuous and discontinuous differentiators*, in "International Journal of Control", 2017, vol. https://doi.org/10.1080/00207179.2017.1393564, p. 1–18

[108] R. USHIROBIRA, A. QUADRAT.*Algebraic estimation of a biased and noisy continuous signal via orthogonal polynomials*, in "55th IEEE Conference on Decision and Control", 2016

[109] M. V. DER PUT, M. SINGER.*Galois Theory of Linear Differential Equations*, Springer, 2002

Project-Team INOCS

Integrated Optimization with Complex Structure

IN PARTNERSHIP WITH:
Ecole Centrale de Lille

Université Libre de Bruxelles

RESEARCH CENTER
Lille - Nord Europe

THEME
Optimization, machine learning and statistical methods

Table of contents

1. Team, Visitors, External Collaborators **137**
2. Overall Objectives **138**
2.1. Introduction 138
2.2. Schedule of tasks 138
3. Research Program **139**
3.1. Introduction 139
3.2. Modeling problems with complex structures 139
3.3. Solving problems with complex structures 140
4. Application Domains **140**
4.1. Energy 140
4.2. Transportation and Logistics 141
4.3. Telecommunications 141
5. Highlights of the Year **141**
5.1.1. Awards 141
5.1.2. Publications & dissemination 141
6. New Software and Platforms **141**
6.1. HappyChic-ApproPick 141
6.2. KEOLIS-MEDIATOUR 142
6.3. PARROT 142
7. New Results **142**
7.1. Large scale complex structure optimization 142
7.2. Bilevel Programming 145
7.3. Robust/Stochastic programming 147
8. Bilateral Contracts and Grants with Industry **148**
8.1. Bilateral Contracts with Industry 148
8.2. Bilateral Grants with Industry 148
8.3. Inria Innovation Lab 148
9. Partnerships and Cooperations **148**
9.1. Regional Initiatives 148
9.1.1. Lille 148
9.1.2. Brussels 148
9.2. National Initiatives 149
9.2.1. ANR 149
9.2.2. F.R.S.-FNRS (Belgium) 149
9.3. International Initiatives 149
9.3.1. Inria International Labs 149
9.3.2. Inria Associate Teams Not Involved in an Inria International Labs 149
9.3.2.1. LOBI 149
9.3.2.2. North-European associated team 150
9.3.3. Inria International Partners 150
9.4. International Research Visitors 151
9.4.1.1. Visiting Professors and Ph.D. students 151
9.4.1.2. Internships 151
10. Dissemination **151**
10.1. Promoting Scientific Activities 151
10.1.1. Scientific Events Organisation 151
10.1.1.1. General Chair, Scientific Chair 151
10.1.1.2. Member of the Organizing Committees 151
10.1.2. Scientific Events Selection 151

10.1.3. Journal 152
10.1.3.1. Member of the Editorial Boards 152
10.1.3.2. Reviewer - Reviewing Activities 152
10.1.4. Invited Talks 152
10.1.5. Leadership within the Scientific Community 152
10.1.6. Scientific Expertise 152
10.1.7. Research Administration 153
10.2. Teaching - Supervision - Juries 153
10.2.1. Teaching 153
10.2.2. Supervision 154
10.2.3. Juries 154
10.3. Popularization 154
11. Bibliography **155**

Project-Team INOCS

Creation of the Team: 2015 May 01, updated into Project-Team: 2018 March 01

Keywords:

Computer Science and Digital Science:

A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.3. - Probabilistic methods
A6.2.6. - Optimization
A9.6. - Decision support

Other Research Topics and Application Domains:

B4. - Energy
B4.3. - Renewable energy production
B4.4. - Energy delivery
B4.5. - Energy consumption
B6. - IT and telecom
B6.3.2. - Network protocols
B7. - Transport and logistics
B7.1. - Traffic management
B7.1.2. - Road traffic
B8.1. - Smart building/home
B8.1.1. - Energy for smart buildings
B8.2. - Connected city
B8.4. - Security and personal assistance

1. Team, Visitors, External Collaborators

Research Scientists

Luce Brotcorne [Team leader, Inria, Researcher, HDR]
Miguel Anjos [Inria, Advanced Research Position, from May 2017]
Markus Sinnl [Inria, Researcher, until Feb 2018]

Faculty Members

Diego Cattaruzza [Ecole centrale de Lille, Associate Professor]
Bernard Fortz [Université libre de Bruxelles, Professor]
Martine Labbé [Université libre de Bruxelles, Professor]
Maxime Ogier [Ecole centrale de Lille, Associate Professor]
Frédéric Semet [Ecole centrale de Lille, Professor]

PhD Students

Mathieu Besancon [Inria, from Sep 2018]
Yaheng Cui [Inria]
Jérôme De Boeck [Université libre de Bruxelles]
Concepción Domínguez Sánchez [Université libre de Bruxelles]
Wenjuan Gu [Ecole centrale de Lille]

Fränk Plein [Université libre de Bruxelles]
Luis Salazar Zendeja [Ecole centrale de Lille, from Nov 2018]
Léonard Von Niederhausern [Inria]
Yuan Yuan [Ecole centrale de Lille]

Post-Doctoral Fellows
Victor Bucarey [Université libre de Bruxelles, from Jun 2018]
Eglantine Camby [Université libre de Bruxelles]
Guillerme Duvillié [Université libre de Bruxelles]
Pablo Escalona Rodriguez [Inria, from Oct 2018]
Arnaud Laurent [Inria, from Sep 2018]
Youcef Magnouche [Inria, until Sep 2018]

Visiting Scientists
Eduardo Alvarez Miranda [Universidad de Talca, until Feb 2018]
Claudia Archetti [Università di Brescia, from Dec 2018]
Sebastian Davila [Universidad de Chile, Jun 2018]
Alejandro Jofre [Universidad de Chile, from Jun 2018 until Jul 2018]
Alfredo Marin [Universidad de Murcia, Mar 2018]
Vladimir Marianov [Pontificia Universidad Católica de Chile, Jun 2018]
Fernando Ordónez [Universidad de Chile, Jun 2018]
Martin Schmidt [Erlangen University, Aug 2018]
Stein Wallace [NHH Norwegian School of Economics, Oct 2018]

Administrative Assistants
Nathalie Bonte [Inria, from Jun 2018]
Corinne Jamroz [Inria, until May 2018]

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. $Y(x) = \text{argmin}\,\{h(x,y)|y \in Y_2, (x,y) \in XY_2\}$ 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 activities in this application domain are supported by bilateral grants/contracts with Colisweb, DHL, HappyChic, INFRABEL, and Kéolis.

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

5.1.1. Awards

- Martim Joyce-Moniz, former INOCS PhD student supervised by Bernard Fortz, won the Best Dissertation Award of the INFORMS Section on Telecommunications and Network Analytics.
- Bernard Fortz, Enrico Gorgone and Dimitri Papadimitriou received the 2017 Glover-Klingman prize for the best paper published in *Networks* (an international journal) [6].
- Wenjuan Gu, together with co-authors Diego Cattaruzza, Maxime Ogier and Frédéric Semet, has been classified finalist for the best article GT2L (Groupe de Travail Transport et Logistique) award with the paper titled *Adaptive large neighborhood search for multicommodity VRP* [49]. The work has been presented during the conference Roadef 2018 held in Lorient, France.

5.1.2. Publications & dissemination

- Luce Brotcorne was the EURO Plenary Speaker at the XIX Latin-Iberoamerican Conference on Operations Research (CLAIO 2018) in Lima, Peru, September 2018 [28].
- Martine Labbé was plenary speaker at the «Journées de l'optimisation» in Montreal, Canada, May 2018 [31].

6. New Software and Platforms

6.1. HappyChic-ApproPick

KEYWORDS: Operational research - Optimization - Java

FUNCTIONAL DESCRIPTION: This software is a prototype developed for the bilateral contract with the company HappyChic. This software is a solver for an integrated warehouse order picking problem with manual picking operations. More precisely, the following problems are solved: (1) the assignment of references to storage positions, based on the iterative solving of minimum cost flow problems, (2) the division of clients orders into several parcels, respecting weight and size constraints, using a dynamic programming algorithm based on the split algorithm, (3) the batching of parcels into trolleys to perform picking tours, using a dynamic programming algorithm based on the split algorithm. The objective function is to minimize the total walking distance. This software is designed to deal with the large-sized industrial instances of HappyChic (considering hundreds of clients, thousands of positions and product references) in a short computation time (few minutes).

- Contact: Maxime Ogier

6.2. KEOLIS-MEDIATOUR

KEYWORDS: Operational research - Mathematical Optimization - Staff scheduling
FUNCTIONAL DESCRIPTION: This software is a prototype developed under a bilateral contract with the company Keolis. This software is a solver which aims to optimize the scheduling of mediation staff More precisely, for each member of the mediation staff working in a public transportation network, MEDIATOUR determines his/her schedule along the day, i.e. when and where he/she is present. Various operational constraints must be taken into account such as the coverage of the network. This software is designed to solve large-scale industrial instances (the subway network of Lille) in short computation times (less than 1 minute).

- Contact: Frédéric Semet

6.3. PARROT

Planning Adapter Performing ReRouting and Optimization of Timing
KEYWORDS: Decision aid - Railway - Scheduling
FUNCTIONAL DESCRIPTION: This is a decision support system addressing the problem of the rescheduling railway schedules on the Belgian network when maintenance operations are planned in the short term (2-3 weeks in advance). The deliverable is a software tool that will take as input: (1) the schedules initially planned for the different trains, (2) the initial routes of the trains, (3) maintenance operations / changes of elements in the form of constraints (unavailable routes etc.). He then provides in output: (1) the new train schedule, (2) the new routing of the fleet. The modifications must respect the constraints corresponding to the operations of maintenance. For example, in some cases it is common to leave at least a few minutes interval between two trains using the same track in the station. This constraint must then be propagated if a maintenance operation delays the arrival of a train. New schedules and routings have to be created following a specific goal. Changes made to schedules and routings must minimize: (1) variations on the time spent at the station, (2) the number of partially canceled trains (additional correspondence (s) or stations that are no longer served), (2) the number of fully canceled trains (no stations served).

- Contact: Martine Labbe

7. New Results

7.1. Large scale complex structure optimization

Formulation and algorithms for last-mile delivery systems:

E-commerce is a thriving market around the world and suits very well the busy lifestyle of today's customers and this growing e-commerce poses a huge challenge for transportation companies, especially in the last mile delivery. We addressed first a fleet composition problem for last-mile delivery service. This problem occurs at a tactical level when the composition of the fleet has to be decided in advance. It is the case for companies that offer last-mile delivery service. Most of them subcontract the transportation part to local carriers and have to decide the day before which vehicles will be needed to cover a partially known demand. We assumed that the distribution area is divided into a limited number of delivery zones and the time horizon into time-slots. The demand is characterized by packages to be transported from pick-up zones to delivery zones given a delivery time slot. First, we introduced an integer programming model which aims to minimize the total delivery cost while ensuring that the demand is covered, the capacity of each vehicle is not violated, the working time for each period is not exceeded and the total working of each delivery respects the social regulations. Then we present a column-generation based approach, which is able to solve real-life instances in reasonable CPU times [33], [32]. Nowadays, the most common last mile delivery service is home delivery. Besides home delivery, companies like Amazon and Fedex, develop locker delivery. When customers shop online, they can choose a nearby locker as a pickup location. In the past years, a new concept called trunk delivery, has been proposed. Here, customers' orders can be delivered to the trunk of their cars. We jointly considered all these delivery possibilities in the same last-mile system and studied the case where the fleet is limited to a single vehicle. We proposed different formulations for the rising optimization problem. We developed problem defined cuts in order to strengthen the formulations and be able to tackle real-size instances. Last we designed and implemented a branch-and-cut algorithm [55], [53].

Large neighborhood algorithm for multi-commodity vehicle routing problem: When delivering fresh fruits and vegetables to catering the multi commodity aspect should be taken into account and deliveries to customers are not made in once, but each commodity can be delivered by a different vehicle as long as the total demand of that commodity is delivered. The problem that arises is the commodity constrained split delivery vehicle routing problem (C-SDVRP). We propose a heuristic based on the adaptive large neighborhood search (ALNS) to efficiently solve medium and large sized instances of the C-SDVRP. We take into account the distinctive features of the C-SDVRP and adapt several local search moves to improve a solution. Moreover, a mathematical programming based operator (MPO) that reassigns commodities to routes is used to improve a new global best solution. Computational experiments have been performed on benchmark instances from the literature. The results assess the efficiency of the algorithm, which can provide a large number of new best-known solutions in short computational times [50].

A matheuristic for the packaging and shipping problem: E-commerce has been continuously growing in the last years to a primary retail market. Recently in France, the threshold of 1 billion of online transactions was overcome. Due to a high demand fluctuation of e-commerce, the workforce sizing for the logistic chain is a challenging problem. Companies have to develop good strategies to have a sustainable workforce size while guaranteeing a high-level service. In this work, we consider the management of the workforce for a warehouse of an e-commerce company. Specifically, we address issues as i) How the workforce at the warehouse can be determined; ii) What is the daily operational production planning; iii) How the demand peaks can be smoothed, and the production maintained ideally constant over the time horizon. To provide answers to these issues, we introduce the Packaging and Shipping Problem (PSP). The PSP looks for a solution approach that jointly determines the workforce over a multi-period horizon and daily operational plans while minimizing the total logistics cost. We consider two strategies that aim to enhance the flexibility of the process and the efficiency of resources use: reassignment and postponement. To tackle the Packaging and Shipping Problem we propose a model, and a three-phase matheuristic. This heuristic is proved to be competitive with respect to the direct solution of the model with a commercial solver on real-life based instances [18].

Heuristic and column generation approaches for the joint order batching and picker routing problem: Picking is the process of retrieving products from the inventory. It is mostly done manually by dedicated employees called pickers and is considered the most expensive of warehouse operations. To reduce the picking cost, customer orders can be grouped into batches that are then collected by traveling the shortest possible distance. We proposed an industrial case study for the HappyChic company where the warehouse has an acyclic layout: pickers are not allowed to backtrack. We developed a two-phase heuristic approach to solve this industrial case [59]. Moreover, we proposed an exponential linear programming formulation to tackle the joint order batching and picker routing problem. Variables, or columns, are related to the picking routes in the warehouse. Computing such routes is in general an intractable routing problem and relates to the well known traveling salesman problem (TSP). Nonetheless, the rectangular warehouse's layouts can be used to efficiently solve the corresponding TSP. Experimented on a publicly available benchmark, the algorithm proves to be very effective. It improves many of the best known solutions and provides very strong lower bounds. This approach is also applied to the HappyChic industrial case to demonstrate its interest for this field of application [41].

Distribution network configuration problems: A distribution network is a system aiming to transfer a certain type of resource from feeders to customers. Feeders are producers of a resource and customers have a certain demand in this resource that must be satisfied. Distribution networks can be represented on graphs and be subject to constraints that limit the number of intermediate nodes between some elements of the network (hop constraints) because of physical constraints. We used layered graphs for hop constrained problems to build extended formulations [21]. Preprocessing techniques allowed to reduce the size of the layered graphs used. The model was studied on the hop-constrained minimum margin problem in an electricity network.

This problem consists of designing a connected electricity distribution network, and to assign customers to electricity feeders at a maximum number of hops H so as to maximize the minimum capacity margin over the feeders to avoid an overload for any feeder. A related theoretical work considers a very special case of hop constrained network design, namely the 2 edge-disjoint 3-paths polyhedron [15].

Switched Ethernet network design problems: We studied models arising in the design of switched Ethernet networks implementing the Multiple Spanning Tree Protocol [23]. In these problems, multiple spanning trees have to be established in a network to route demands partitioned into virtual local access networks. Different mixed-integer formulations for the problem have been proposed and compared, both theoretically and computationally.

Delay management in public transportation: The Delay Management Problem arises in Public Transportation networks, and is characterized by the necessity of connections between different vehicles. The attractiveness of Public Transportation networks is strongly related to the reliability of connections, which can be missed when delays or other unpredictable events occur. Given a single initial delay at one node of the network, the Delay Management Problem is to determine which vehicles have to wait for the delayed ones, with the aim of minimizing the dissatisfaction of the passengers. We derived strengthened mixed integer linear programming formulations and new families of valid inequalities for that problem. The implementation of branch-and-cut methods and tests on a benchmark of instances taken from real networks show the potential of the proposed formulations and cuts [20].

Discrete ordered median problem: The discrete ordered median problem consists in locating p facilities in order to minimize an ordered weighted sum of distances between clients and closest open facility. We formulate this problem as a set partitioning problem using an exponential number of variables. Each variable corresponds to a set of demand points allocated to the same facility with the information of the sorting position of their corresponding costs. We develop a column generation approach to solve the continuous relaxation of this model. Then, we apply a branch-price-and-cut algorithm to solve small to large sized instances of DOMP in competitive computational time [62].

Genome wide association studies: We studied the Polymorphic Alu Insertion Recognition Problem (PAIRP). Alu (Arthrobacter luteus) forms a major component of repetitive DNA and are frequently encountered during the genotyping of individuals. The basic approach to find Alus consists of (i) aligning sequence reads from a set of individual(s) with respect to a reference genome and (ii) comparing the possible Alu insertion induced by the alignment with the Alu insertions positions already known for the reference genome. The sequence genome of the reference individual is known and will be highly similar, but not identical, to the genome of the individual(s) being sequenced. Hence, at some locations they will diverge. Some of this divergence is due to the insertion of Alu polymorphisms. Detecting Alus has a central role in the field of Genetic Wide Association Studies because basic elements are a common source of mutation in humans. We investigated the PAIRP relationship with the the Clique Partitioning of Interval Graphs (CPIG). Our results [12], [26] provide insights of the complexity of the problem, a characterization of its combinatorial structure and an exact approach based on Integer Linear Programming to exactly solve the correspond instances.

A branch-and-cut algorithm for the maximum k-balanced subgraph of a signed graph: A signed graph is k-balanced if its vertex set can be partitioned into at most k sets in such a way that positive edges are found only within the sets and negative edges go between sets. The maximum k-balanced subgraph problem is the problem of finding a subgraph of G that is k-balanced and maximum according to the number of vertices. This problem has applications in clustering problems appearing in collaborative vs conflicting environments. We provide a representatives formulation for the problem and present a partial description of the associated polytope, including the introduction of strengthening families of valid inequalities. A branch-and-cut algorithm is described for finding an optimal solution to the problem. An ILS metaheuristic is implemented for providing primal bounds for this exact method and a branching rule strategy is proposed for the representatives formulation. Computational experiments, carried out over a set of random instances and on a set of instances from an application, show the effectiveness of the valid inequalities and strategies adopted in this work [22].

Feature selection in support vector machine: This work focuses on support vector machine (SVM) with feature selection. A MILP formulation is proposed for the problem. The choice of suitable features to construct the separating hyperplanes has been modelled in this formulation by including a budget constraint that sets in advance a limit on the number of features to be used in the classification process. We propose both an exact and a heuristic procedure to solve this formulation in an efficient way. Finally, the validation of the model is done by checking it with some well-known data sets and comparing it with classical classification methods [25].

7.2. Bilevel Programming

Pricing problems in energy management: Power systems face higher flexibility requirements from generation to consumption due to the increasing penetration of non-controllable distributed renewable energy. In this context, demand side management aims at reducing excessive load fluctuation and match the price of energy to their real cost for the grid. Pricing models for demand side management methods are traditionally used to control electricity demand. First, we proposed bilevel pricing models to explore 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. Moreover, we assumed that the smart grid optimizes the usage of a renewable energy generation source and a storage capacity. Results over a rolling horizon were obtained [14], [28], [36]. Next, we considered four types of actors: furnishers sell electricity, local agents trade and consume energy, aggregators trade energy and provide energy to end-users, who consume it. This gives rise to three levels of optimization. The interaction between aggregators and their end-users is modelled with a bilevel program, and so is the interaction between furnishers, and local agents and aggregators. Since solving bilevel programs is difficult in itself, solving trilevel programs requires particular care. We proposed three possible approaches, two of them relying on a characterization of the intermediary optimization level [35]. Finally, Time and-Level-of-Use is a recently proposed energy pricing scheme, designed for the residential sector and providing suppliers with robust guarantee on the consumption. We formulate the supplier decision as a bilevel, bi-objective problem optimizing for both financial loss and guarantee. A decomposition method is proposed, related to the optimal value transformation. It allows for the computation of an exact solution by finding possible Pareto optimal candidate solutions and then eliminating dominated ones. Numerical results on experimental residential power consumption data show the method effectively finds the optimal candidate solutions while optimizing costs only or incorporating risk aversion at the lower-level [37].

Unit commitment under market equilibrium constraints: Traditional (deterministic) models for the Unit Commitment problem (UC) assume that the net demand for each period is perfectly known in advance, or in more recent and more realistic approaches, that a set of possible demand scenarios is known (leading to stochastic or robust optimization problems). However, in practice, the demand is dictated by the amounts that can be sold by the producer at given prices on the day-ahead market. We modeled and solved the UC problem with a second level of decisions ensuring that the produced quantities are cleared at market equilibrium. In its simplest form, we are faced to a bilevel optimization problem where the first level is a MIP and the second level linear. As a first approach to the problem, we assumed that demand curves and offers of competitors in the market are known to the operator. Following the classical approach for these models, we turned the problem into a single-level program by rewriting and linearizing the first-order optimality conditions of the second level. In recent work, this approach was extended to include network capacities effects and a decoupling of prices in different zones [45], [46], [47], [48].

Market regulation: We proposed a bilevel programming model to study a problem of market regulation through government intervention. One of the main characteristics of the problem is that the government monopolizes the raw material in one industry, and competes in another industry with private firms for the production of commodities. Under this scheme, the government controls a state-owned firm to balance the market; that is, to minimize the difference between the produced and demanded commodities. On the other hand, a regulatory organism that coordinates private firms aims to maximize the total profit by deciding the amount of raw material bought from the state-owned firm. Two equivalent single-level reformulations are proposed to solve the problem. Additionally, three heuristic algorithms are designed to obtain good-quality solutions with low computational effort. Extensive computational experimentation is carried out to measure the efficiency of the proposed solution methodologies. A case study based on the Mexican petrochemical industry is presented. Additional instances generated from the case study are considered to validate the robustness of the proposed heuristic algorithms [66].

Rank pricing: One of the main concerns in management and economic planning is to sell the right product to the right customer for the right price. Companies in retail and manufacturing employ pricing strategies to maximize their revenues. The Rank Pricing Problem considers a unit-demand model with unlimited supply and uniform budgets in which customers have a rank-buying behavior. Under these assumptions, the problem is first analyzed from the perspective of bilevel pricing models and formulated as a non linear bilevel program with multiple independent followers. We also present a direct non linear single level formulation. Two different linearizations of the models are carried out and two families of valid inequalities are obtained which, embedded in the formulations by implementing a branch-and-cut algorithm, allow us to tighten the upper bound given by the linear relaxation of the models. We show the efficiency of the formulations, the branch-and-cut algorithms and some preprocessing through extensive computational experiments [16].

Bilevel minimum spanning tree problem: Consider a graph whose edge set is partitioned into a set of red edges and a set of blue edges, and assume that red edges are weighted and contain a spanning tree of G. Then, the Bilevel Minimum Spanning Tree Problem (BMSTP) consists in pricing (i.e., weighting) the blue edges in such a way that the total weight of the blue edges selected in a minimum spanning tree of the resulting graph is maximized. We propose different mathematical formulations for the BMSTP based on the properties of the Minimum Spanning Tree Problem and the bilevel optimization. We establish a theoretical and empirical comparison between these new formulations and we also provide reinforcements that together with a proper formulation are able to solve medium to big size instances [65].

Bilevel programming models for location problems: First, we addressed a multi-product location problem in which a retail firm has several malls with a known location. A particular product comes in p types. Each mall has a limited capacity for products to be sold at that location, so the firm has to choose what products to sold at what mall. Furthermore, the firm can apply discrete levels of discount on the products/ The objective of the firm is to find what products to sell at which mall, with what level of discount, so that its profit is maximized. Consumers are located in points of the region. Each consumer has a different set of acceptable products, and will purchase one of these, or none if it is not convenient for her. Consumers maximize their utility. The agents (firm and consumers) play a Stackelberg game, in which the firm is the leader and the customers the follower. Once the firm decides the products to sell at each mall and the possible discounts, consumers purchase (or not) one of their acceptable products wherever their utility is maximized. We model the problem using bilevel formulations, which are compared on known instances from the literature [43]. Second we studied a location problem of controversial facilities. On the one hand, a leader chooses among a number of fixed potential locations which ones to establish. On the second hand, one or several followers who, once the leader location facilities have been set, choose their location points in a continuous framework. The leader's goal is to maximize some proxy to the weighted distance to the follower's location points, while the follower(s) aim is to locate his location points as close as possible to the leader ones. We develop the bilevel location model for one follower and for any polyhedral distance, and we extend it for several followers and any so-called p-norm. We prove the NP-hardness of the problem and propose different mixed integer linear programming formulations. Moreover, we develop alternative Benders decomposition algorithms for the problem. Finally, we report some computational results comparing the formulations and the Benders decompositions on a set of instances [63].

Stackelberg games: First 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 identify 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 [61].

Second, we formulate a Stackelberg Security game that coordinates resources in a border patrol problem. In this security domain, resources from different precincts have to be paired to conduct patrols in the border due to logistic constraints. Given this structure the set of pure defender strategies is of exponential size. We describe the set of mixed strategies using a polynomial number of variables but exponentially many constraints that come from the matching polytope. We then include this description in a mixed integer formulation to compute the Strong Stackelberg Equilibrium efficiently with a branch and cut scheme. Since the optimal patrol solution is a probability distribution over the set of exponential size, we also introduce an efficient sampling method that can be used to deploy the security resources every shift. Our computational results evaluate the efficiency of the branch and cut scheme developed and the accuracy of the sampling method. We show the applicability of the methodology by solving a real world border patrol problem [58].

7.3. Robust/Stochastic programming

Locating stations in a one-way electric car sharing system under demand uncertainty: We focused in [60] on a problem of locating recharging stations in one-way station based electric car sharing systems which operate under demand uncertainty. We modeled this problem as a mixed integer stochastic program and develop a Benders decomposition algorithm based on this formulation. We integrated a stabilization procedure to our algorithm and conduct a large-scale experimental study on our methods. To conduct the computational experiments, we developed a demand forecasting method allowing to generate many demand scenarios. The method was applied to real data from Manhattan taxi trips.

Integrated shift scheduling and load assignment optimization for attended home delivery: We studied an integrated shift scheduling and load assignment optimization problem for attended home delivery. The proposed approach is divided into two phases, each one corresponding to a different planning level: tactical and operational. In the tactical planning, a daily master plan is generated for each courier. This master plan defines the working shifts, the origin-destination pairs to visit, and the number of client requests to serve. In the operational planning, delivery orders are allocated to couriers in real-time. The stochastic and dynamic nature of client orders is included in the tactical and operational decision levels, respectively. For the tactical level, we developed and implemented a multi-cut L-shaped algorithm. Experimental results demonstrate that our approach provides robust tactical solutions that easily accommodate to fluctuations in client orders, preventing additional costs related to the underutilization of couriers and to the use of external couriers to satisfy all delivery requests, when compared to an approach using the mean demand value. Moreover, these results also indicate that the failure to incorporate robust tactical solutions in the operational planning results in infeasible operational plans that are inadmissible regarding the couriers' working time (e.g., minimum and maximum number of working hours) and work regulations (e.g., allocation of consecutive working hours to the couriers).

Bookings in the European gas market: Characterization of feasibility and computational complexity results: As a consequence of the liberalisation of the European gas market in the last decades, gas trading and transport have been decoupled. At the core of this decoupling are so-called bookings and nominations. Bookings are special long-term capacity right contracts that guarantee that a specified amount of gas can be supplied or withdrawn at certain entry or exit nodes of the network. These supplies and withdrawals are nominated at the day-ahead. These bookings then need to be feasible, i.e., every nomination that complies with the given bookings can be transported. While checking the feasibility of a nomination can typically be done by solving a mixed-integer nonlinear feasibility problem, the verification of feasibility of a set of bookings is much harder. We consider the question of how to verify the feasibility of given bookings for a number of special cases. For our physics model we impose a steady-state potential-based flow model and disregard controllable network elements. We derive a characterization of feasible bookings, which is then used to show that the problem is in coNP for the general case but can be solved in polynomial time for linear potential-based flow models. Moreover, we present a dynamic programming approach for deciding the feasibility of a booking in tree-shaped networks even for nonlinear flow models [56].

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

Fluxys (2016-2018). Study of optimization problems arising in the management of gas networks.

HappyChic (2018). Study and implementation of optimization methods for problems arising in the warehouse management context.

Keolis (2018). Study and implementation of optimization methods for problems arising in the management of mediation officers in public transportation.

Utocat (2018). Study optimization problems arising in the blockchain.

8.2. Bilateral Grants with Industry

Design and Pricing of Electricity Services in a Competitive Environment within the Gaspard Monge Research Program (PGMO) funded by the Fondation Mathématiques Jacques Hadamard. EDF is the industrial partner (2015-2018).

Robust Energy Offering under Market Equilibrium Constraints within the Gaspard Monge Research Program (PGMO) funded by the Fondation Mathématiques Jacques Hadamard. EDF is the industrial partner (2017-2019).

8.3. Inria Innovation Lab

COLINOCS is an Inria Innovation Lab between Colisweb, a start-up company addressing last-mile delivery and INOCS, which was created at the end of 2016. This collaboration roots back to 2015, when a bilateral contract was devoted to optimization problems arising in courier scheduling. The main objective of this Innovation Lab is to model and solve optimization problems related to revenue management, transport mutualization, a better visibility on their activities for the couriers. See also: https://www.inria.fr/centre/lille/actualites/inria-innovation-lab-colinocs-entre-colisweb-et-l-equipe-inocs.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Lille

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).

9.1.2. Brussels

ValueBugs is a citizen participatory research project, funded by INNOVIRIS (2018-2020). The objective of ValueBugs is to collectively develop a method for decentralized insect production in cities while enhancing the value of food waste on a small scale. In practical terms, peelings are consumed by insect larvae that have reached the end of their development and offer many promising outlets: feed for hens, farmed fish, pets... and much more! This new, totally innovative sector will be a new tool to be put in the hands of every citizen: we must therefore imagine it collectively.

9.2. National Initiatives

9.2.1. 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 - 2019). 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.

9.2.2. F.R.S.-FNRS (Belgium)

Bilevel optimization is a branch of mathematical optimization that deals with problems whose constraints embed an auxiliary optimization problem. The F.R.S.-FNRS research project “bilevel optimization” (2018-2019) will study such bilevel problems with bilinear objectives and simple second level problems. Each follower chooses one strategy in a given fixed set of limited size. Two classes of such problems will be studied: Pricing Problems and Stackelberg Security Games.

In pricing problems, prices for products must be determined to maximize the revenue of a leader given specific behaviors of customers (followers). More precisely, we will consider the single minded pricing problem and the rank pricing problem.

In Stackelberg games, mixed strategies to cover targets,must be determined in order to maximize the defender expected payoff given that attackers (followers) attack targets that maximize their own payoffs.

9.3. International Initiatives

9.3.1. Inria International Labs

Inria Chile

Associate Team involved in the International Lab:

9.3.1.1. BIPLOS

Title: BIlevel Problems in LOgistics and Security

International Partner (Institution - Laboratory - Researcher):

Universidad de Chile (Chile) - Instituto Sistemas Complejos de Ingeieria (ISCI) - Ordonez Fernando

Start year: 2017

See also: https://project.inria.fr/biplos/

This projet is devoted to bilevel optimisation problems with application in the security and logistics domains. Stackelberg games, including one defender and several followers, and competitive location problems will be considered. Mixed integer linear optimisation models and efficient algorithms to solve them will be developed.

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

9.3.2.1. LOBI

Title: Learning within Bilevel Optimization

International Partner (Institution - Laboratory - Researcher):

Polytechnique Montréal (Canada) - Institut de Valorisation des Données (IVADO) - Gilles Savard

Start year: 2018

See also: https://team.inria.fr/lobi/

The interplay between optimization and machine learning is one of the most important developments in modern computational science. Simultaneously there is a tremendous increase in the availability of large quantities of data in a multitude of applications, and a growing interest in exploiting the information that this data can provide to improve decision-making. Given the importance of big data in business analytics, its explicit integration into an optimization process is a challenge with high potential impact. The innovative project is concerned with the interconnection between machine learning approaches and a particular branch of optimization called bilevel optimization in this "big data" context. More precisely, we will focus on the development of new approaches integrating machine learning within bilevel optimization (LOBI: "Learning au sein de l'Optimisation BIniveau") for two important practical applications, the pricing problem in revenue management and the energy ressource aggregation problem in smart grids. The applications arise from current industry collaborations of the teams involved, and will serve as testbeds to demonstrate the potential impact of the proposed approach.

9.3.2.2. North-European associated team

Title: Physical-internet services for city logistics

International Partner (Institution - Laboratory - Researcher):

Norwegian School of Economics - Stein Wallace

Start year: 2017

In this project, we consider an urban logistic terminal and new logistics services which could be developed according to a Physical Internet approach. The main objective is to evaluate the services using optimization models created within the project. We are developing optimization models to identify win-win cooperation between carriers based on supply and demand. We aim to explore how to include stochasticity in the description of the supplies and demands, as well as travel times, and to what extent the plans within a day can improve by such knowledge. The second task is to develop solution algorithms for these models. These are real scientific challenges as we are facing stochastic mixed integer problems.

9.3.3. Inria International Partners

9.3.3.1. 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 (CIRRELT), Montreal, Canada.

Department of Industrial Engineering, Universidad de Talca, Curicó, Chile.

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 Electrical and Information Engineering, University of Padova, Italy.

Department of Mathematics, University of Aveiro, Portugal.

Department of Statistics and Operations Research, University of Lisbon, Portugal.

Instituto de Matemáticas, University of Seville, Spain.

Departamento de Estadística e Investigación Operativa, Universidad de Murcia, Spain.

Dipartimento di Matematica, Universita degli studi di Padova, Italy.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

9.4.1.1. Visiting Professors and Ph.D. students

Claudia Archetti, Professor at Universita de Brescia, December 2018.

Stein Wallace, Professor at NHH Norwegian School of Economics, October 2018.

Martin Schmidt, Professor at Erlangen University, August 2018.

Alejandro Jofre, Professor at Universidad de Chile, from June 2018 until July 2018.

Sebastián Dávila, Ph.D. student at Universidad de Chile, June 2018.

Vladimir Marianov, Professor at Pontificia Universidad Católica de Chile, June 2018.

Fernando Ordónez, Professor at Universidad de Chile, June 2018.

Alfredo Marin, Professor at Universidad de Murcia, March 2018.

Eduardo Alvarez Miranda, Professor at Universidad de Talca, January - February 2018.

9.4.1.2. Internships

Brou Koua, Esatic, Abidjan, Cote d'Ivoire, December 2018 - January 2019.

Lilian Lopez Vera, Autonomous University of Nuevo León, Monterrey, Mexico February-June 2018.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

- IWOBIP 2018, 2nd International Workshop on Bilevel Programming, Lille, France, June 2018: Luce Brotcorne, Martine Labbé.
- Operations Research 2018, Brussels, Belgium, September 2018: Bernard Fortz.

10.1.1.2. Member of the Organizing Committees

7th Winter School on Network Optimization, Estoril, Portugal, January 2018: Bernard Fortz.

International Symposium on Mathematical Programming (ISMP), Bordeaux, July 2018: Luce Brotcorne, Bernard Fortz.

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

ORBEL 2018, Liège, Belgium, January 2018: Bernard Fortz

6th INFORMS Transportation Science and Logistics Society Workshop, January, 2018: Luce Brotcorne, Frédéric Semet

ROADEF2018 - 18ème Conférence de la Société Française de Recherche Opérationnelle et d'Aide à la Décision, Metz, France, February 2018: Luce Brotcorne, Bernard Fortz, Frédéric Semet

ISCO 2018: International Symposium on Combinatorial Optimization, Marrakesh, Morocco, April, 2018: Martine Labbé

Odysseus 2018, International Workshop on Freight Transportation, Cagliari, Italy, June 2018: Martine Labbé, Frédéric Semet

Matheuristics 2018, Tours, France, June 2018: Martine Labbé

EURO/ALIO International Conference 2018 on Applied Combinatorial Optimization, Bologna, June 2018: Martine Labbé

International Symposium on Mathematical Programming (ISMP), Bordeaux, July 2018: Martine Labbé (Program Committee), Frédéric Semet (Scientific Committee)

EURO 2018, European Conference of Operational Research Societies, Valencia, Spain, July 2018: Luce Brotcorne, Bernard Fortz

INFORMS Annual Meeting, Phoenix, United States, November 2018: Luce Brotcorne (cluster chair), Bernard Fortz

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

EURO Journal on Computational Optimization: Martine Labbé - Editor in chief, Bernard Fortz - Editor.

Computers and Operations Research: Luce Brotcorne - Associate editor.

INFORMS Journal on Computing: Bernard Fortz - Associate editor.

International Transactions in Operations Research: Bernard Fortz, Martine Labbé - Associate editors.

Transportation Science: Martine Labbé - Member of the Advisory Board.

10.1.3.2. Reviewer - Reviewing Activities

Annals of Operations Research, Applied Computing and Informatics, Central European Journal of Operations Research, Computers & Operations Research, Computational Optimization and Applications, Discrete Applied Mathematics, EURO Journal on Transportation and Logistics, European Journal of Operational Research, IISE Transactions, INFORMS Journal on Computing, International Journal of Management Science and Engineering Management, Mathematical Programming Computation, Networks, Omega, Operations Research, Optimization and Engineering, RAIRO - Operations Research, Transportation Science: Luce Brotcorne, Diego Cattaruzza, Bernard Fortz, Martine Labbé, Maxime Ogier, Frédéric Semet.

10.1.4. Invited Talks

XIX Latin-Iberoamerican Conference on Operations Research (CLAIO 2018), Lima, Peru, September 2018: Luce Brotcorne, EURO Plenary speaker [28].

Journées de l'optimisation, Montreal, Canada, May 2018 : Martine Labbé, Plenary Speaker [31].

10.1.5. Leadership within the Scientific Community

EURO Working Group "Pricing and Revenue Management": Luce Brotcorne - coordinator.

EURO Working Group "European Network Optimization Group (ENOG)": Bernard Fortz - coordinator.

EURO Working Group "Vehicle routing and logistics optimization (VEROLOG)": Frédéric Semet - Member of the board.

INFORMS Women in OR/MS : Luce Brotcorne - International liaison.

SIAG/Optimization Prize committee: Martine Labbé - Chair.

ORBEL (Belgian Operations Research Society): Bernard Fortz - Member of the board of administration and treasurer.

ORBEL representative for EURO and IFORS: Bernard Fortz.

CNRS GdR 3002: Operations Research: Frédéric Semet - Member of the steering committee.

10.1.6. Scientific Expertise

Scientific orientation committee of the Interuniversity Centre on Entreprise Networks, Transportation and Logistics (CIRRELT), Canada: Bernard Fortz, Frédéric Semet - Members.

Centro de Matemática, Aplicaçoes Fundamentais e Investigaçao Operacional, University of Lisbon: Martine Labbé - Member.

DFG Review Panel “Mathematics” for Clusters of Excellence, 2018: Martine Labbé - Member.

Scientific committee of France-Netherlands Exchange Program: Luce Brotcorne - Member.

Evaluation committee for Inria/MITACS Exchange Program: Luce Brotcorne - Member.

Evaluation committee COST GTRI: Luce Brotcorne - Member.

President of the FRIA PE1 - jury 1: Bernard Fortz - Chair.

Scientific board of PICOM competitiveness cluster: Frédéric Semet - Member.

Agence Nationale de la Recherche (ANR): Luce Brotcorne, Frédéric Semet - Reviewer.

Fond de Recherche Nature et Technologie du Québec: Frédéric Semet - Reviewer.

Research Council of Norway: Frederic Semet - Reviewer.

10.1.7. Research Administration

Committee for the Technological Development (CDT): Luce Brotcorne - Member.

CRIStAL: Frédéric Semet - Deputy-director.

Scientific council of Centrale Lille: Frédéric Semet - Elected member.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: Bernard Fortz, Recherche Opérationnelle et Applications, 30hrs, M1, Université de Mons (campus Charleroi), Belgique.

Master: Bernard Fortz, Continuous Optimization, 24hrs, M1 & M2, Université libre de Bruxelles, Belgique.

Master: Martine Labbé, Computer science seminar, 12hrs, M2, Université libre de Bruxelles, Belgique.

Master: Frédéric Semet, Non-linear Optimization, 30hrs, M2, Centrale Lille.

Master: Frédéric Semet, Operations Research, 28hrs, M2, Centrale Lille.

Master: Luce Brotcorne, Optimisation, 14hrs, M1, Polytech Lille.

Master: Luce Brotcorne, Recherche opétrationnelle, 16hrs, M1 apprentissage, Polytech Lille.

Master: Luce Brotcorne, Diego Cattaruzza, Maxime Ogier, Frédéric Semet, Numerical Analysis and Optimization, 132hrs, M1, Centrale Lille.

Master: Diego Cattaruzza, Maxime Ogier, Object-Oriented Programming, 48hrs, M1, Centrale Lille.

Master: Diego Cattaruzza, Maxime Ogier, Operations Research, 16hrs, M1, Centrale Lille.

Master: Frédéric Semet, Large-scale optimization methods, 24hrs, M1, Centrale Lille.

Licence: Diego Cattaruzza, Maxime Ogier, Object-Oriented Programming, 36hrs, L3, Centrale Lille.

Licence: Frédéric Semet, Advanced programming and Complexity, 24hrs, L3, Centrale Lille.

Licence: Diego Cattaruzza, Maxime Ogier, Object-Oriented Programming, 40hrs, L2, Centrale Lille.

Licence: Diego Cattaruzza, Web Technologies and Multimedia, 32hrs, L2, Centrale Lille.

Licence: Bernard Fortz, Algorithmique 1, 12hrs, L1, Université libre de Bruxelles, Belgique.

Licence: Bernard Fortz, Algorithmique 2, 24hrs, L2, Université libre de Bruxelles, Belgique.

Licence: Bernard Fortz, Algorithmique et Recherche Opérationnelle, 24hrs, L3, Université libre de Bruxelles, Belgique.

Licence: Martine Labbé, Projets d'informatique 3 transdisciplinaire, 12hrs, L3, Université libre de Bruxelles, Belgique.

10.2.2. Supervision

PhD: Luciano Porretta, Models and methods for the study of genetic associations, Université libre de Bruxelles, January 2018, Bernard Fortz.

PhD in progress: Jérôme De Boeck, Optimization problems in energy, from October 2015, Bernard Fortz.

PhD in progress: Mathieu Besançon, Approche bi-niveau de réponse à la demande dans les réseaux électriques intelligents, from September 2018, Miguel Anjos, Luce Brotcorne, Frédéric Semet.

PhD in progress: Yaheng Cui, Models and methods for decentralized decision in logistics networks, from Oct 2016, Luce Brotcorne, Eric Ballot.

PhD in progress: Concepción Domínguez Sánchez, Mixed integer linear models and algorithms for pricing problems, from October 2017, Martine Labbé.

PhD in progress: Wenjuan Gu, Location routing for short and local fresh food supply 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: Fränk Plein, Models and methods for the robust verification of booked capacities in gas networks in a decentralized setting, from October 2017, Martine Labbé.

PhD in progress: Luis Alberto Salazar Zendeja, Formulations and resolution methods for network interdiction problems, from November 2018, Diego Cattaruzza, Martine Labbé, Frédéric Semet.

PhD in progress: Yuan Yuan, Vehicle routing problems with synchronization for city logistics, from Oct 2016, Diego Cattaruzza, Frédéric Semet.

10.2.3. Juries

PhD: "Design of reliable aerospace system architecture", University of Edinburgh. Lukas Matthias Schaefer, Bernard Fortz - Reviewer.

PhD: "Fast and scalable optimization for segment routing", Université catholique de Louvain. Renaud Hartert, Bernard Fortz - Examiner.

PhD: "Network design under uncertainty and demand elasticity", Concordia University. Carlos Armando Zetina, Bernard Fortz - External examiner.

Habilitation: "Optimization for complex problems in air and rail transport", Université de Lille 1. Paola Pellegrini, Bernard Fortz - Reviewer.

Habilitation: "A journey through optimization: from global to discrete optimization and back", Université de Lorraine. Bernardetta Addis, Bernard Fortz - Reviewer, Martine labbé - Examiner.

PhD: "Design of optimal routes for vehicles and drivers", Universidad de La Laguna. Bencomo Domínguez-Martín, Martine Labbé - Reviewer.

PhD: "A comparative study of labeling algorithms within the branch-and-price framework for vehicle routing with time windows", Université de Liège. Stefano Michelini, Martine Labbé - Examiner.

PhD: "Mathematical programming models and algorithms for offshore wind park design", Technical University of Denmark. Martina Fischetti, Martine Labbé - External examiner.

Habilitation: "Linear formulations and exact algorithms for combinatorial optimization", Sorbonne Université. Pierre Fouilhoux, Martine Labbé - Examiner.

PhD: "Tactical planning on freight transport networks: service design and pricing", Université de Liège. Christine Tawfik, Luce Brotcorne - Reviewer.

PhD: "A tropical geometry and discrete convexity approach to bilevel programming", Université de Paris-Saclay. Jean-Bernard Eytard, Luce Brotcorne - Reviewer.

PhD: "Revenue management for transport service provides in physical internet: Freight carriers as case", Université de Recherche Paris Sciences et Lettres. Bin Qiao, Luce Brotcorne - Reviewer.

10.3. Popularization

F. Semet, Club Logistique et Transport, CCI St Quentin, November 2018.

11. Bibliography

Major publications by the team in recent years

[1] 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

[2] 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^o 1, p. 13-26 [*DOI :* 10.1287/IJOC.1110.0472], http://joc.journal.informs.org/content/25/1/13.abstract

[3] L. BROTCORNE, F. CIRINEI, P. MARCOTTE, G. SAVARD.*An exact algorithm for the network pricing problem*, in "Discrete Optimization", 2011, vol. 8, n^o 2, p. 246–258 [*DOI :* 10.1016/J.DISOPT.2010.09.003]

[4] L. BROTCORNE, M. LABBÉ, P. MARCOTTE, G. SAVARD.*Joint design and pricing on a network*, in "Operation Research", 2008, vol. 56, p. 1104-1115 [*DOI :* 10.1287/OPRE.1080.0617]

[5] D. CATTARUZZA, N. ABSI, D. FEILLET.*The Multi-Trip Vehicle Routing Problem with Time Windows and Release Dates*, in "Transportation Science", 2016, vol. 50, n^o 2, p. 676-693, https://doi.org/10.1287/trsc.2015.0608

[6] 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^o 1, p. 110–123, http://dx.doi.org/10.1002/net.21721

[7] 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

[8] B. FORTZ, M. THORUP.*Internet Traffic Engineering by Optimizing OSPF Weights*, in "Proc. 19th IEEE Conf. on Computer Communications (INFOCOM)", 2000, p. 519-528

[9] S. GARCÍA, M. LABBÉ, A. MARÍN.*Solving large p-median problems with a radius formulation*, in "INFORMS Journal on Computing", 2011, vol. 23, n^o 4, p. 546–556

[10] 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, vol. 50, n^o 4, p. 1286-1299, https://doi.org/10.1287/trsc.2016.0692

[11] M. LABBÉ, P. MARCOTTE, G. SAVARD.*A bilevel model of taxation and its application to optimal highway pricing*, in "Management science", 1998, vol. 44, n^o 12-part-1, p. 1608–1622

Publications of the year

Doctoral Dissertations and Habilitation Theses

[12] L. PORRETTA.*Models and methods in genome wide association studies*, Université libre de Bruxelles, January 2018, https://hal.inria.fr/tel-01944087

Articles in International Peer-Reviewed Journal

[13] R. ABSIL, E. CAMBY, A. HERTZ, H. MELOT.*A sharp lower bound on the number of non-equivalent colorings of graphs of order n and maximum degree n - 3*, in "Discrete Applied Mathematics", January 2018, vol. 234, p. 3-11, https://hal.inria.fr/hal-01944274

[14] E. ALEKSEEVA, L. BROTCORNE, S. LEPAUL, A. MONTMEAT.*A bilevel approach to optimize electricity prices*, in "Yugoslav Journal of Operations Research", 2018 [*DOI :* 10.2298/YJOR171115002A], https://hal.inria.fr/hal-01971679

[15] Q. BOTTON, B. FORTZ, L. GOUVEIA.*The 2 edge-disjoint 3-paths polyhedron*, in "Annals of Telecommunications - annales des télécommunications", 2018, vol. 73, n^o 1-2, p. 29–36 [*DOI :* 10.1007/S12243-017-0615-2], https://hal.inria.fr/hal-01673313

[16] H. CALVETE, C. DOMÍNGUEZ, C. GALÉ, M. LABBÉ, A. MARÍN.*The Rank Pricing Problem: models and branch-and-cut algorithms*, in "Computers and Operations Research", 2018, https://hal.inria.fr/hal-01782958

[17] E. CAMBY, G. CAPOROSSI, M. PAIVA, M. E. V. SEGATTO.*Expected distance based on random walks*, in "Journal of Mathematical Chemistry", February 2018, vol. 56, n^o 2, p. 618-629, https://hal.inria.fr/hal-01944246

[18] D. CATTARUZZA, L. BROTCORNE, F. SEMET, B. TOUNSI.*A three-phase matheuristic for the Packaging and Shipping Problem*, in "Applied Mathematical Modelling", August 2018, vol. 64, p. 713-732, https://hal.archives-ouvertes.fr/hal-01959341

[19] M. A. D. R. S. CRUZ, H. R. D. O. ROCHA, M. H. M. PAIVA, M. E. V. SEGATTO, E. CAMBY, G. CAPOROSSI.*An algorithm for cost optimization of PMU and communication infrastructure in WAMS*, in "International Journal of Electrical Power and Energy Systems", March 2019, vol. 106, p. 96-104 [*DOI :* 10.1016/J.IJEPES.2018.09.020], https://hal.inria.fr/hal-01944258

[20] V. DAL SASSO, L. DE GIOVANNI, M. LABBÉ.*Strengthened Formulations and Valid Inequalities for Single Delay Management in Public Transportation*, in "Transportation Science", 2018, https://hal.inria.fr/hal-01925451

[21] J. DE BOECK, B. FORTZ.*Extended formulation for hop constrained distribution network configuration problems*, in "European Journal of Operational Research", March 2018, vol. 265, n^o 2, p. 488 - 502 [*DOI :* 10.1016/J.EJOR.2017.08.017], https://hal.inria.fr/hal-01665624

[22] R. FIGUEIREDO, Y. FROTA, M. LABBÉ.*A branch-and-cut algorithm for the maximum k-balanced subgraph of a signed graph*, in "Discrete Applied Mathematics", 2018, https://hal.inria.fr/hal-01937015

[23] B. FORTZ, L. GOUVEIA, M. JOYCE-MONIZ.*Optimal design of switched Ethernet networks implementing the Multiple Spanning Tree Protocol*, in "Discrete Applied Mathematics", January 2018, vol. 234, p. 114 - 130 [*DOI :* 10.1016/J.DAM.2016.07.015], https://hal.inria.fr/hal-01665614

[24] B. FORTZ, D. PAPADIMITRIOU, M. RESENDE.*Preface: Recent advances in telecommunications networks planning and operation*, in "Networks", March 2018, vol. 71, n^o 2, p. 95-96, https://hal.inria.fr/hal-01943965

[25] M. LABBÉ, L. I. MARTÍNEZ-MERINO, A. M. RODRÍGUEZ-CHÍA.*Mixed Integer Linear Programming for Feature Selection in Support Vector Machine*, in "Discrete Applied Mathematics", 2018, https://hal.inria.fr/hal-01924379

[26] L. PORRETTA, D. CATANZARO, B. HALLDÓRSSON, B. FORTZ.*A Branch & Price algorithm for the minimum cost clique cover problem in max-point tolerance graphs*, in "4OR: A Quarterly Journal of Operations Research", June 2018, https://hal.inria.fr/hal-01943982

[27] M. RESTREPO, F. SEMET, T. POCREAU.*Integrated Shift Scheduling and Load Assignment Optimization for Attended Home Delivery*, in "Transportation Science", 2018, https://hal.inria.fr/hal-01963916

Invited Conferences

[28] L. BROTCORNE.*How to integrate customer's behaviour within pricing*, in "CLAIO 2018 - XIX Latin-Iberoamerican Conference on Operations Research", Lima, Peru, September 2018, https://hal.inria.fr/hal-01971738

[29] M. LABBÉ.*Bilevel optimisation, pricing and Stackelberg problems*, in "Winter School on Network Optimization", Estoril, Portugal, January 2018, https://hal.inria.fr/hal-01958961

[30] M. LABBÉ.*Bilevel Programming: Pricing Problems and Stackelberg Games*, in "IWR Doctoral School on Optimization", Heidelberg, Germany, October 2018, https://hal.inria.fr/hal-01959004

[31] M. LABBÉ.*Stackelberg games and Bilevel Bilinear Optimization Problems*, in "Journées de l'optimisation", Montréal, Canada, May 2018, https://hal.inria.fr/hal-01958984

[32] F. SEMET, E. ALEKSEEVA, L. BROTCORNE, Y. MAGNOUCHE, E. SOUFFLET.*Fleet composition for last-mile delivery service*, in "EURO 2018 - 29th European Conference on Operational Research", Valencia, Spain, July 2018, https://hal.inria.fr/hal-01964556

[33] F. SEMET.*Some logistics and transportation optimization problems in the E-commerce industry*, in "Journées du CIRRELT - 2018", Québec, Canada, May 2018, https://hal.inria.fr/hal-01964218

International Conferences with Proceedings

[34] S. BELIERES, N. JOZEFOWIEZ, F. SEMET.*A Graph Reduction Heuristic For Supply Chain Transportation Plan Optimization*, in "ODYSSEUS 2018 - Seventh International Workshop on Freight Transportation and Logistics", Cagliari, Italy, June 2018, 4, https://hal.laas.fr/hal-01876531

[35] L. BROTCORNE, D. AUSSEL, L. VON NIEDERHÄUSERN.*A Trilevel Energy Market Model*, in "ISMP 2018 - 23rd International Symposium on Mathematical Programming", Bordeaux, France, July 2018, https://hal.inria.fr/hal-01971754

[36] L. BROTCORNE, S. LEPAUL, L. VON NIEDERHÄUSERN.*Pricing in the energy field*, in "European Conference on Operational Research", Valencia, Spain, July 2018, https://hal.inria.fr/hal-01971715

[37] L. BROTCORNE, A. F. MIGUEL, M. BESANÇON, J. A. GOMEZ-HERRERA.*Demand Response Planner for Building Districts*, in "Informs Annual Meeting", Phoenix, United States, November 2018, https://hal.inria.fr/hal-01971730

[38] S. MICHEL, L. BROTCORNE, D. CATTARUZZA, N. MITTON, F. SEMET.*A heuristic approach for the computation of individual trajectories of a fleet of robots under connectivity constraints*, in "Congrès annuel de la société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF)", Lorient, France, February 2018, https://hal.inria.fr/hal-01704705

[39] L. VON NIEDERHÄUSERN, L. BROTCORNE, S. LEPAUL.*A Rolling Horizon Method for a Bilevel Demand Side Management Problem*, in "PGMO Days 2018", Paris, France, November 2018, https://hal.inria.fr/hal-01971767

[40] Y. YUAN, D. CATTARUZZA, M. OGIER, F. SEMET.*The last mile delivery problem*, in "ROUTE 2018 - International Workshop on Vehicle Routing, Intermodal Transportation and Related Areas", Snekkersten, Denmark, May 2018, https://hal.inria.fr/hal-01964234

Conferences without Proceedings

[41] O. BRIANT, H. CAMBAZARD, D. CATTARUZZA, N. CATUSSE, A.-L. LADIER, M. OGIER.*A column generation based approach for the joint order batching and picker routing problem*, in "ROADEF 2018 - 19ème Congrès Annuel de la Société Française de Recherche Opérationnelle et d'Aide à la Décision", Lorient, France, February 2018, p. 1-2, https://hal.archives-ouvertes.fr/hal-01765113

[42] L. BROTCORNE, Y. MAGNOUCHE, F. SEMET, E. ALEKSEEVA.*A Revenue Management Approach for Attended Home Delivery*, in "The 6th INFORMS Transportation Science and Logistics Society Workshop", Hong-Kong, China, January 2018, https://hal.inria.fr/hal-01964679

[43] S. DAVILA, M. LABBÉ, V. MARIANOV, F. ORDÓÑEZ, F. SEMET.*Bilevel programming models for multi-product location problems*, in "IWOBIP'18 - 2nd International Workshop on Bilevel Programming", Lille, France, June 2018, https://hal.inria.fr/hal-01964681

[44] J. DE BOECK, M. LABBÉ, E. MARCOTTE, P. MARCOTTE, G. SAVARD.*Dynamic programming approach for bidding problems on day-ahead markets*, in "ISMP 2018 - 23rd International Symposium on Mathematical Programming", Bordeaux, France, July 2018, https://hal.inria.fr/hal-01958995

[45] B. FORTZ, L. BROTCORNE, F. D'ANDREAGIOVANNI, J. DE BOECK.*Unit Commitment under Market Equilibrium Constraints*, in "Journées de l'Optimisation 2018", Montréal, Canada, May 2018, https://hal.inria.fr/hal-01944151

[46] B. FORTZ, L. BROTCORNE, F. D'ANDREAGIOVANNI, J. DE BOECK.*Unit Commitment under Market Equilibrium Constraints*, in "ISMP2018 - 23rd International Symposium on Mathematical Programming", Bordeaux, France, July 2018, https://hal.inria.fr/hal-01944201

[47] B. FORTZ, L. BROTCORNE, F. D'ANDREAGIOVANNI, J. DE BOECK.*Unit Commitment under Market Equilibrium Constraints*, in "IWOBIP'18 - 2nd International Workshop on Bilevel Programming", Lille, France, June 2018, https://hal.inria.fr/hal-01944168

[48] B. FORTZ, L. BROTCORNE, F. D'ANDREAGIOVANNI, J. DE BOECK.*Unit Commitment under Market Equilibrium Constraints*, in "ROADEF 2018 - 19ème congré annuel de le société Française de Recherche Opérationnelle et d'Aide à la Décision", Lorient, France, February 2018, https://hal.inria.fr/hal-01944100

[49] W. GU, D. CATTARUZZA, M. OGIER, F. SEMET.*Adaptive large neighborhood search for multicommodity VRP*, in "ROADEF 2018 - 19ème Congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision", Lorient, France, February 2018, https://hal.archives-ouvertes.fr/hal-01940685

[50] W. GU, D. CATTARUZZA, M. OGIER, F. SEMET.*Adaptive large neighborhood search for multicommodity VRP*, in "Odysseus 2018 - the Seventh International Workshop on Freight Transportation and Logistics", Cagliari, Italy, June 2018, https://hal.archives-ouvertes.fr/hal-01951948

[51] W. GU, D. CATTARUZZA, M. OGIER, F. SEMET.*Adaptive large neighborhood search for the commodity constrained split delivery VRP*, in "OR 2018 - International Conference on Operations Research", Brussels, Belgium, September 2018, https://hal.archives-ouvertes.fr/hal-01948464

[52] P. NOLZ, N. ABSI, D. CATTARUZZA, D. FEILLET.*Two-echelon distribution with city hub capacity management*, in "Odysseus 2018 - 7th International Workshop on Freight Transportation and Logistics", Cagliari, Italy, June 2018, https://hal.archives-ouvertes.fr/hal-01966725

[53] Y. YUAN, D. CATTARUZZA, M. OGIER, F. SEMET.*A branch-and-cut algorithm for the generalized traveling salesman problem with time windows*, in "OR 2018 - International Conference on Operations Research", Brussels, Belgium, September 2018, https://hal.archives-ouvertes.fr/hal-01941124

[54] Y. YUAN, D. CATTARUZZA, M. OGIER, F. SEMET.*A branch-and-cut algorithm for the generalized traveling salesman problem with time windows*, in "ROADEF 2018 - 19ème Congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision", Lorient, France, February 2018, https://hal.archives-ouvertes.fr/hal-01941090

[55] Y. YUAN, D. CATTARUZZA, M. OGIER, F. SEMET.*Last mile delivery problem: the one-vehicle case*, in "Odysseus 2018 - 7th Workshop on Freight Transportation and Logistics", Cagliari, Italy, June 2018, https://hal.archives-ouvertes.fr/hal-01951934

Research Reports

[56] M. LABBÉ, F. PLEIN, M. SCHMIDT.*Bookings in the European Gas Market: Characterisation of Feasibility and Computational Complexity Results*, Université Libre de Bruxelles (U.L.B.), Belgium, December 2018, https://hal.inria.fr/hal-01954262

Other Publications

[57] M. BESANÇON, L. BROTCORNE, M. F. ANJOS, J. A. GOMEZ-HERRERA.*Increase in power demand guarantee: a bilevel approach*, May 2018, Optimization days, https://hal.inria.fr/hal-01971748

[58] V. BUCAREY, C. CASORRÁN, M. LABBÉ, F. ORDÓÑEZ, O. FIGUEROA.*Coordinated defender strategies for border patrols*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01917782

[59] M. BUÉ, D. CATTARUZZA, M. OGIER, F. SEMET.*An integrated order batching and picker routing problem*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01849980

[60] H. CALIK, B. FORTZ.*A Benders decomposition method for locating stations in a one-way electric car sharing system under demand uncertainty*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01962059

[61] C. CASORRÁN, B. FORTZ, M. LABBÉ, F. ORDÓÑEZ.*A study of general and security Stackelberg game formulations*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01917798

[62] S. DELEPLANQUE, M. LABBÉ, D. PONCE LOPEZ, J. PUERTO.*A Branch-Price-and-Cut Procedure for the Discrete Ordered Median Problem*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01954865

[63] M. LABBÉ, M. LEAL, J. PUERTO.*New models for the location of controversial facilities: A bilevel programming approach*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01933601

[64] M. LABBÉ, P. MARCOTTE.*Bilevel Network Design*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01937014

[65] M. LABBÉ, M. POZO, J. PUERTO.*Computational comparisons of different formulations for the Bilevel Minimum Spanning Tree Problem*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01937013

[66] H. MARAVILLO, J.-F. CAMACHO-VALLEJO, J. PUERTO, M. LABBÉ.*A market regulation bilevel problem: a case study of the Mexican petrochemical industry*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01922518

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

1. Team, Visitors, External Collaborators **166**
2. Overall Objectives **166**
2.1. Overall Objectives 166
2.2. Presentation 166
3. Research Program **167**
3.1. Background 167
3.2. Querying Heterogeneous Linked Data 167
3.3. Managing Dynamic Linked Data 168
3.4. Linking Graphs 169
4. Application Domains **170**
4.1. Linked Data Integration 170
4.2. Data Cleaning 170
4.3. Real Time Complex Event Processing 170
5. Highlights of the Year **170**
6. New Software and Platforms **171**
6.1. ShEx validator 171
6.2. gMark 171
6.3. SmartHal 172
6.4. QuiXPath 172
6.5. X-FUN 172
7. New Results **172**
7.1. Querying Heterogeneous Linked Data 172
7.1.1. Data Integration 172
7.1.2. Schema Validation 172
7.2. Managing Dynamic Linked Data 173
7.2.1. Complex Event Processing 173
7.2.2. Transformations 173
7.3. Foundations of AI 173
7.3.1. Knowledge Compilation 173
7.3.2. Aggregation and Enumeration for Graphs 174
8. Bilateral Contracts and Grants with Industry **174**
9. Partnerships and Cooperations **174**
9.1. Regional Initiatives 174
9.2. National Initiatives 175
9.3. European Initiatives 176
9.4. International Initiatives 176
9.5. International Research Visitors 176
9.5.1. Visits of International Scientists 176
9.5.2. Visits to International Teams 177
10. Dissemination **177**
10.1. Promoting Scientific Activities 177
10.1.1. Scientific Events Organisation 177
10.1.2. Scientific Events Selection 177
10.1.2.1. Chair of Conference Program Committees 177
10.1.2.2. Member of the Conference Program Committees 177
10.1.3. Journal 177
10.1.4. Invited Talks 177
10.1.5. Scientific Expertise 177
10.1.6. Research Administration 178

10.2. Teaching - Supervision - Juries 178
10.2.1. Teaching 178
10.2.2. Supervision 178
10.2.3. Juries 178
10.3. Popularization 178
10.3.1. Internal action 178
10.3.2. Creation of media or tools for science outreach 179
11. Bibliography **179**

Project-Team LINKS

Creation of the Team: 2013 January 01, updated into Project-Team: 2016 June 01

Keywords:

Computer Science and Digital Science:

A2.1. - Programming Languages
A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.6. - Concurrent programming
A2.4. - Formal method for verification, reliability, certification
A2.4.1. - Analysis
A2.4.2. - Model-checking
A2.4.3. - Proofs
A3.1. - Data
A3.1.1. - Modeling, representation
A3.1.2. - Data management, quering and storage
A3.1.3. - Distributed data
A3.1.4. - Uncertain data
A3.1.5. - Control access, privacy
A3.1.6. - Query optimization
A3.1.7. - Open data
A3.1.8. - Big data (production, storage, transfer)
A3.1.9. - Database
A3.2.1. - Knowledge bases
A3.2.2. - Knowledge extraction, cleaning
A3.2.3. - Inference
A3.2.4. - Semantic Web
A4.7. - Access control
A4.8. - Privacy-enhancing technologies
A7. - Theory of computation
A7.2. - Logic in Computer Science
A9.1. - Knowledge
A9.2. - Machine learning
A9.7. - AI algorithmics
A9.8. - Reasoning

Other Research Topics and Application Domains:

B6.1. - Software industry
B6.3.1. - Web
B6.3.4. - Social Networks
B6.5. - Information systems
B9.5.1. - Computer science
B9.5.6. - Data science
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientist

Joachim Niehren [Team leader, Inria, Senior Researcher, HDR]

Faculty Members

Iovka Boneva [Université de Lille, Associate Professor]
Florent Capelli [Université de Lille, Associate Professor]
Aurélien Lemay [Université de Lille, Associate Professor, HDR]
Charles Paperman [Université de Lille, Associate Professor]
Sylvain Salvati [Université de Lille, Professor, HDR]
Slawomir Staworko [Université de Lille, Associate Professor, HDR]
Sophie Tison [Université de Lille, Professor, HDR]

Technical Staff

Jeremie Dusart [Inria]

PhD Students

Nicolas Crosetti [Inria, from Oct 2018]
Dimitri Gallois [Université de Lille]
Paul Gallot [Inria]
Jose Martin Lozano [Université de Lille]
Momar Sakho [Inria]

Post-Doctoral Fellows

Nicolas Bacquey [Inria, until Aug 2018]
Bruno Guillon [Inria, from Nov 2018]

Visiting Scientists

Rustam Azimov [Saint Petersburg University, from Aug 2018 until Nov 2018]
Herminio Garcia Gonzalez [Oviedo University, from Sep 2018 until Dec 2018]

Administrative Assistant

Nathalie Bonte [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 summarize 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 parallelization (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

Containment for RDF Schemas

The ShEx language for defining RDF schemas was proposed and developed earlier by the Links team in cooperation with the W3C. Slawek Staworko et al. now studied the containment problem for ShEx schemas for RDF documents. They showed at **PODS** [7] – the best database theory conference – that the problem is decidable, but `co-NEXP`-hard. This is joined work with P. Wieczorek from the University of Wrazlaw.

Foundations of AI: Knowledge Compilation

Florent Capelli et al. showed at **STACS** [15] – a top conferences in theoretical computer science – a new knowledge compilation procedure for quantified boolean formulas allowing to decide the satisfiability quantified boolean formulas with bounded tree width in polynomial time. This can be applied in particular to first-order database queries with quantifiers. This is joined work with S. Mengel from the CNRS in Lens.

Foundations of AI: Constrained Topologial Sort

Charles Paperman et al. showed at **ICALP** [8] – a top conferences in theoretical computer science – how to compute efficiently topological sorts of graphs under regular constraints. The problem was initially introduced in the context of preferential query answer for uncertain databases, where on usually wants to sort the query answers by some preferences, that are known only partially. It becomes then crucial to look for total orders on the answer set satisfying regular constraints that specify the preferences. Finding such an order for regular constraints was know to be infeasible in general. In this article, a class of regular constraints is identified for which this problem becomes tractable. A (partial) decidable dichotomy theorem is proven drawing the frontier between the kind of constraints which are feasible from those which are not. This is joined work with A. Amarilli from Telecom Paristech.

6. New Software and Platforms

6.1. ShEx validator

Validation of Shape Expression schemas
KEYWORDS: Data management - RDF
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.

RELEASE FUNCTIONAL DESCRIPTION: ShExJava now uses the Commons RDF API and so support RDF4J, Jena, JSON-LD-Java, OWL API and Apache Clerezza. It can parse ShEx schema in the ShEcC, ShEJ, ShExR formats and can serialize a schema in ShExJ.

To validate data against a ShExSchema using ShExJava, you have two different algorithms: - the refine algorithm: compute once and for all the typing for the whole graph - the recursive algorithm: compute only the typing required to answer a validate(node,ShapeLabel) call and forget the results.

- Contact: Iovka Boneva
- URL: http://shexjava.lille.inria.fr/

6.2. gMark

gMark: schema-driven graph and query generation
KEYWORDS: Semantic Web - Data base
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 - ability to generate queries with a desired selectivity

- Contact: Aurélien Lemay
- URL: https://github.com/graphMark/gmark

6.3. SmartHal

KEYWORD: Bibliography

FUNCTIONAL DESCRIPTION: SmartHal is a better tool for querying the HAL bibliography database, while is based on Haltool queries. The idea is that a Haltool query returns an XML document that can be queried further. In order to do so, SmartHal provides a new query language. Its queries are conjunctions of Haltool queries (for a list of laboratories or authors) with expressive Boolean queries by which answers of Haltool queries can be refined. These Boolean refinement queries are automatically translated to XQuery and executed by Saxon. A java application for extraction from the command line is available. On top of this, we have build a tool for producing the citation lists for the evaluation report of the LIFL, which can be easily adapter to other Labs.

- Contact: Joachim Niehren
- URL: http://smarthal.lille.inria.fr/

6.4. QuiXPath

KEYWORDS: XML - NoSQL - Data stream

SCIENTIFIC DESCRIPTION: The QuiXPath tools supports a very large fragment of XPath 3.0. The QuiXPath library provides a compiler from QuiXPath to FXP, which is a library for querying XML streams with a fragment of temporal logic.

FUNCTIONAL DESCRIPTION: QuiXPath is a streaming implementation of XPath 3.0. It can query large XML files without loading the entire file in main memory, while selecting nodes as early as possible.

- Contact: Joachim Niehren
- URL: https://project.inria.fr/quix-tool-suite/

6.5. X-FUN

KEYWORDS: Programming language - Compilers - Functional programming - Transformation - XML

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: Joachim Niehren and Pavel Labath
- Contact: Joachim Niehren

7. New Results

7.1. Querying Heterogeneous Linked Data

7.1.1. Data Integration

The PhD project of Lozano on relational to RDF data integration is progressing under the direction of Boneva, and Staworko. At **AMW** [9] they studied the *relational to RDF data exchange problem*. They focus in particular on a preliminary analysis of the consistency problem for relational to RDF data exchange with target ShEx schema.

7.1.2. Schema Validation

Shape Expression Language 2.0 (ShEx) is a language to describe the vocabulary and the structure of an RDF graph. It is base on the notion of shapes, a typing system supporting algebraic operations, recursive references to other shapes or Boolean combination.

In their **PODS** paper [7], Staworko studied the *containment problem* for ShEx (in cooperation with Wieczorek from Wrazlaw). Containement is a classical subject for schema-related issue in database theory. The authors proved that it is decidable for ShEx-schema, but with a untractable complexity (co-NEXP-hard). They also carefuly craft restriction of ShEx schema to design tractable-but-still-signifiant fragments.

7.2. Managing Dynamic Linked Data

7.2.1. Complex Event Processing

Complex event processing requires to answer queries on streams of complex events, i.e., nested words or equivalently linearizations of data trees, but also to produce dynamically evolving data structures as output.

The topic of the PhD project of M. Sakho supervised by Niehren and Boneva is to generalize algorithms for querying streams to hyperstreams. These are collections of linked streams as naturally produced as intermediate results of complex events processing. Hyperstreams are incomplete descriptions of relational structures, so they can queried similarly to incomplete databases, for which the notion of a certain query answer is most appropriate.

In a paper published at **RP** [13], they studied certain query answering for hyperstreams with simple events. Such hyperstreams can be identified with compressed string patterns. They proved that the certain query answering for regular queries on compressed string patterns is PSPACE-complete, independently of whether the finite automata defining the regular queries are assumed deterministic or not, and independently of whether compression is permitted or not. They also showed that the problem is in PTIME when restricted to *linear* string patterns (possibly with compression) and to deterministic finite automata.

In a paper published at **LATA** [6], the studied certain query answering on hyperstreams of complex events. Such hyperstreams can be modeled by compressed tree pattern with context variables. They showed that certain query answering for regular queries on compressed tree pattern with context variables is EXP-complete, independently of whether the tree automata defining the regular queries are assumed deterministic or not, and independently of whether compression is permitted or not. They also showed that the problem is in PTIME when restricted to *linear* tree patterns (possibly with compression) and to deterministic tree automata.

7.2.2. Transformations

In his PhD project – belonging to the ANR Colis– Gallot with his supervisors Salvati and Lemay presented higher order tree transducers which extend macro tree transducers. Moreover they obtained nice properties such as the closure of the transducers under composition. Algorithms to compute such compositions are proposed. Those algorithms perform partial evaluation and are guided by semantic interpretations over finite domains.

Another virtue of higher-order transducers is that their *linear* syntactic restriction make them equivalent to logically defined MSO transductions. One of the composition algorithm proposed preserves the linearity. Furthermore, we have also showed that we can decrease the order of linear transducer (i.e. the complexity of the functions it handles) when this one is larger than 4.

These results are unpublished paper for now.

7.3. Foundations of AI

Various problems of databases and knowledge bases are closely related to foundational problems in artificial intelligence, since they are rooted in logic or graph theory.

7.3.1. Knowledge Compilation

Many problems in Artificial Intelligence boil down to the exploration of the solution set (called the models) of logical formulas. Such an exploration can be finding one model of the formula, counting the number of models or enumerating them all. However, even for simple quantifier-free formulas, those explorations are known be untractable (NP-hard).

Knowledge compilation encompasses methods that aim to change the representation of the set of models in order to get tractable algorithms for (some of) those tasks. A big computational cost is paid during the compilation time but then replying to queries become tractable on the new representation. More generally, the core of Knowledge compilation is the study of the trade-off between the size of the representation and the easiness of queries. This subject is of interest for both Artificial Intelligence and Database communities.

At **STACS** [15], Capelli, in cooperation with Mengel from CRIL (Lens), studied knowledge compilation techiques for quantified Boolean formulas. Deciding the existence of models for such formulas is known to climb arbitrarly high the polynomial time hierarchy. The authors provide an efficient compilation procedure for formulas having a *bounded tree-width* generalizing results from SAT solving.

7.3.2. Aggregation and Enumeration for Graphs

Aggregation and enumeration are not relevant for answer sets of database queries but equally for any kinds of sets, most typically defined by combinatoric problems on graphs.

In a paper published at **ICALP** [8], Paperman proposed (in cooperation with Amarilli from Telecom Paristech) to study the problem of finding so called *topological sort* satisfying constraints provided by regular expressions. Searching topological sort happens typically in situations where an order is *uncertain*. For instance, in relational database where users provides a partial preference order, or in concurrent and distributed programming where some tasks can be executed in an arbitrary order. A classical task in *preferential query answering* is to find a topological sort satisfying some global constrained. Typically, to find a total order satisfying all (or most) of the customers. The paper provides and proves sufficient conditions on the *shape of the constraints* to make the problem tractable (P-time) as well as sufficient condition to make the problem NP-hard. They also prove a complete dichotomy for an adapted and well chosen version of the constrained topological sort problem.

In an article in **JCSS** [2], Capelli (with Bergougnoux and Kanté from Bordeaux and Clérmont-Ferrand) propose an algorithm for counting the number of *transversal* in some *hypergraphs*. Here, a hypergraph is a collection of sets – called *hyperedges* over a *ground set* and a traversal is a subset intersecting all hyperedges. In full generality, counting the number of minimal traversals in a hypergraph is a hard problem: it is known to be $\#P$-complete. They proved that under the assumptions of β-acyclicity, it is possible to count all the minimal traversals can be done in polynomial times.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Grants with Industry

Posos. A. Lemay is directing an internship of a master student (*projet de fin d'étude*) in cooperation with the POSOS company from Amiens. The goal of this collaboration is to work on efficient schema for a large pharmaceutical Knowledge Base.

Strapdata. C. Paperman is actively collaborating with the Strapdata company on efficient distributed graph database using an Apache novel technology to query distributed graph *Gremlin* that could benefit of the main product of Strapdata: Elassandra as a *database backend*.

9. Partnerships and Cooperations

9.1. Regional Initiatives

- Links is member of the CPER Data (2016-19).
- Lozano's PhD project (2016-19) is co-funded by the Region Nord-Pas de Calais.
- Sakho's PhD project is co-funded by the Region Nord-Pas de Calais.

- Gillot's PhD project (2017-20) is co-funded by the Region Nord-Pas de Calais.
- Crosetti's PhD project (2018-21) is co-funded by the Region Haut de France. This is joined work with J. Ramon from the Inria project Magnet.

9.2. National Initiatives

ANR Aggreg (2014-19): Aggregated Queries.

- Participants: J. Niehren [correspondent], P. Bourhis, A. Lemay, A. Boiret, F. Capelli.
- The coordinator is J. Niehren and the partners 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, N. Bacquey, P. Gallot, S. Salvati.
- 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: J. Niehren, M. Sakho, N. Crosetti, F. Capelli.
- 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: J. Niehren, S. Salvati, A. Lemay, N. Bacquey, 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.

ANR Bravas (2017-22):

- Participants: S. Salvati [correspondent]
- Scientific Partners: The coordinator is Jérôme Leroux from LaBRI, University of Bordeaux. The other partner is LSV, ENS Cachan.
- Objective: The goal of the BraVAS project is to develop a new and powerful approach to decide the reachability problems for Vector Addition Systems (VAS) extensions and to analyze their complexity. The ambition here is to crack with a single hammer (ideals over well-orders) several long-lasting open problems that have all been identified as a barrier in different areas, but that are in fact closely related when seen as reachability.

9.3. European Initiatives

Oxford: A exchange project with the computer science lab of the University of Oxford is funded by the University of Lille via the Cristal Lab. Links' member produced many common publications over the years with Oxford. Links' contact is Paperman.

Wrazlaw: Staworko has regular exchange with the University of Wrazlaw. This has led to a publication at **PODS** [7] together with P. Wieczorek.

Saint Petersburg: Salvati and Niehren started a cooperation with the University of Saint Petersburg, via a 3 months visit of R. Azimov in 2018.

Oviedo: Boneva started a cooperation with the University of Oviedo, via a 3 months visit of H. Garcia Gonzalez in 2018.

9.4. International Initiatives

9.4.1. Informal International Partners

Santiago de Chile: S. Staworko started a collaboration with C. Riveros from the Pontificia Universidad Catolica de Chile in 2018.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Several researchers has visited us:

- Filip Mazowiecki, a researcher from Warsaw University and currently in post-doctorate in Bordeaux to work with Charles Paperman.
- Rustam Azimov, a Russian PhD students from Saint Petersburg State University, to collaborate with Sylvain Salvati and Joachim Niehren.
- Michaël Cadilhac, a researcher from Oxford University to work with Charles Paperman.
- Cristian Riveros, an Assistant Professor at the Department of Computer Science at the Pontificia Universidad Catolica de Chile.
- Henning Fernau, Professor at Universität Trier and Andreas Maletti, Professor at Universität Leipzig, visited us during the HDR defense of Aurelien Lemay.

9.5.1.1. Internships

- Nicolas Crosetti started an internship supervised by Florent Capelli, Joachim Niehren and Jan Ramon. His internship has evolved into the preparation of a PhD thesis.
- Chen Huan, from Centrale Lille, has done an internship under the supervision of Sylvain Salvati and Joachim Niehren.

9.5.2. Visits to International Teams

- Charles Paperman visited Michaël Cadilhac form the verification team of the University of Oxford.
- Joachim Niehren got invited by Hilal Zaid for a visit at the American University of Palestine in August 2018.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

- F. Capelli: workshop organisation of *Graph and Constraints* (27/08) within the conference Constraint Programming (CP) 2018, Lille.
- F. Capelli: organisation of annual meeting of GT ALGA (Groupe de Travail Automata, Logic, Games, Algebra of CNRS) the 15th and 16th of October at Lille.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

- J. Niehren was is was chair of the Program Committee of WPTE 2018.
- J. Niehren was is was co-chair of the Program Committee of WPTE 2019.

10.1.2.2. Member of the Conference Program Committees

- F. Capelli: member of Program Committee of International Joint Conference on Artificial Intelligence (IJCAI) 2018.
- F. Capelli: member of Program Committee of workshop Quantified Boolean Formulas (QBF) within FLoC conference (Federated Logic Conference).
- F. Capelli: member of Program Committee of workshop Graph and Constraints, 2018.
- S. Tison: member of Program Committee of RAIRO ITA, 2018.
- J. Niehren: member of the Program Committee of LATA 2019.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

J. Niehren is editor of Fundamenta Informaticæ.

S. Salvati is managing editor of JoLLI (Journal for Logic, Language and Information).

S. Tison is in the editorial committee of RAIRO-ITA (Theoretical Informatics and Applications).

10.1.4. Invited Talks

- B. Guillon gave invited talks at Mid-term Meeting of ANR Delta in Bordeaux in December 2018.
- A. Lemay gave invited talks at Mid-term Meeting of ANR Delta in Bordeaux in December 2018.
- Joachim Niehren gave an invited talk at the American University of Palestine in August 2018.
- F. Capelli get invited to talk at seminars of CRIL at Lens, LACL in Créteil and VALDA in Paris.

10.1.5. Scientific Expertise

- S. Tison: member of coordinator of i-SIte ULNE, about innovation and relationship with social economical world.
- S. Tison: Head of CITC-Eurarfid.
- J. Niehren is member of the board of the committee of project-teams of Inria Lille.

10.1.6. Research Administration

- F. Capelli: Co-organizer of *Groupe de Travail* of CNRS IMIA (Informatique Mathématique Intelligence Artificielle)

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

I. Boneva gives a Master 1 semester on Algorithms for databases.

C. Paperman is pedagogical responsible for Master MIASHS “parcours” WebAnalyste.

A. Lemay is pedagogical responsible for Computer Science and Numeric correspondent for UFR LEA.

J. Niehren was teaching the course “Foundations of Databases” as part of the masters 2 Mocad on Information Extraction at the University of Lille.

S. Salvati is pedagogical responsible of Master Miage FA, Lille.

S. Salvati organized the research label for Computer Science Bachelor, Lille.

S. Tison is pedagogical responsible of first year ACT Master, Lille.

10.2.2. Supervision

HdR : Aurélien Lemay, *Machine Learning Techniques for Semistructured Data*, Université de Lille, Fri 16th Nov.

PhD in progress: N. Crosetti. Privacy Risks of Aggregates in Data Centric-Workflows. Supervised by Capelli, Niehren, Ramon (Team MAGNET) and Tison.

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.

PhD in progress: P. Gallot. On safety of data transformations. Started on October 2017. Supervised by Lemay and Salvati.

10.2.3. Juries

S. Tison: Vice-Présidente du jury Agrégation de Mathématiques (co-pillote option D- Informatique).

S. Tison: Jury de Thèse Lucien Mousin.

S. Tison: PhD Committee of Narjes Jomaa.

F. Capelli: PhD Committee of Mikaël Monet.

J. Niehren: Habilitation Committee of Aurélien Lemay.

10.3. Popularization

10.3.1. Internal action

General Assembly of Inria Lille Niehren presented Links work on data and knowledge bases on the Dynamic Semantic Crosswords demonstration during a general assembly of Inria Lille in July 2018.

10.3.2. Creation of media or tools for science outreach

Dynamic Semantic Crosswords Bacquey's demonstration system on dynamic semantic crosswords is presented in the new showroom of Inria Lille in the new building Place. The demo generates dynamically crosswords while streaming Twitter feeds, depending on a semantic topic specified by the user. The specification can be given by a list of hashtags, and in the future by a XPath 3.0 query, that can be executed on streams by using Links QuiXPath tool. This illustrates the work on complex event processing by Niehren and his students during the last years.

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] A. LEMAY.*Machine Learning Techniques for Semistructured Data*, Université de Lille, November 2018, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01929944

Articles in International Peer-Reviewed Journal

[2] B. BERGOUGNOUX, F. CAPELLI, M. M. KANTÉ.*Counting Minimal Transversals of β-Acyclic Hypergraphs*, in "Journal of Computer and System Sciences", November 2018, https://arxiv.org/abs/1808.05017 [*DOI :* 10.1016/J.JCSS.2018.10.002], https://hal.inria.fr/hal-01923090

[3] A. BOIRET, V. HUGOT, J. NIEHREN, R. TREINEN.*Logics for Unordered Trees with Data Constraints*, in "Journal of Computer and System Sciences", December 2018, 40, https://hal.inria.fr/hal-01176763

[4] F. CAPELLI, Y. STROZECKI.*Incremental delay enumeration: Space and time*, in "Discrete Applied Mathematics", August 2018, https://arxiv.org/abs/1703.01928 [*DOI :* 10.1016/J.DAM.2018.06.038], https://hal.inria.fr/hal-01923091

[5] L. DAVIAUD, C. PAPERMAN.*Classes of languages generated by the Kleene star of a word*, in "Information and Computation", October 2018, vol. 262, n^o Part 1, p. 90-109, https://hal.archives-ouvertes.fr/hal-01943493

International Conferences with Proceedings

[6] M. SAKHO, I. BONEVA, J. NIEHREN.*Regular Matching and Inclusion on Compressed Tree Patterns with Context Variables*, in "13th International Conference on Language and Automata Theory and Applications (LATA)", Saint Petersburg, Russia, Springer, March 2019, https://hal.inria.fr/hal-01926011

[7] S. STAWORKO, P. WIECZOREK.*Containment of Shape Expression Schemas for RDF*, in "SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems (PODS)", Amsterdam, Netherlands, June 2019, https://hal.inria.fr/hal-01959143

Conferences without Proceedings

[8] A. AMARILLI, C. PAPERMAN.*Topological Sorting with Regular Constraints*, in "ICALP 2018 - 45th International Colloquium on Automata, Languages, and Programming", Prague, Czech Republic, July 2018, https://hal.archives-ouvertes.fr/hal-01950909

[9] I. BONEVA, J. LOZANO, S. STAWORKO.*Relational to RDF Data Exchange in Presence of a Shape Expression Schema*, in "AMW 2018 - 12th Alberto Mendelzon International Workshop on Foundations of Data Management", Cali, Colombia, May 2018, p. 1-16, https://hal.archives-ouvertes.fr/hal-01775199

[10] I. BONEVA, J. NIEHREN, M. SAKHO.*Certain Query Answering on Compressed String Patterns: From Streams to Hyperstreams*, in "RP 2018 - 12th International Conference on Reachability Problems", Marseille, France, September 2018, https://hal.archives-ouvertes.fr/hal-01609498

Research Reports

[11] S. SALVATI.*On is an n-MCFL*, Université de Lille, Inria, CRIStAL CNRS, April 2018, https://hal.archives-ouvertes.fr/hal-01771670

Other Publications

[12] I. BONEVA, J. NIEHREN, M. SAKHO.*Approximating Certain Query Answering on Hyperstreams*, June 2018, Technical report, https://hal.inria.fr/hal-01811835

[13] I. BONEVA, J. NIEHREN, M. SAKHO.*Certain Query Answering on Compressed String Patterns: From Streams to Hyperstreams (long version)*, July 2018, working paper or preprint, https://hal.inria.fr/hal-01846016

[14] F. CAPELLI, N. CROSETTI, J. NIEHREN, J. RAMON.*Dependency Weighted Aggregation on Factorized Databases*, January 2019, https://arxiv.org/abs/1901.03633 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01981553

[15] F. CAPELLI, S. MENGEL.*Knowledge Compilation, Width and Quantification*, July 2018, https://arxiv.org/abs/1807.04263 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01836402

[16] F. CAPELLI, Y. STROZECKI.*Enumerating models of DNF faster: breaking the dependency on the formula size*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01891483

Team LOKI

technology and knowledge for interaction

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

1. Team, Visitors, External Collaborators **185**
2. Overall Objectives **186**
2.1. Introduction 186
2.2. Knowledge for Interaction 186
2.2.1. Computers as tools 186
2.2.2. Empowering tools 186
2.2.3. Transparent tools 187
2.3. Technology for Interaction 187
2.3.1. “Designeering Interaction” 187
2.3.2. Computers vs Interactive Systems 187
2.3.3. The Interaction Machine 188
3. Research Program **188**
3.1. Introduction 188
3.2. Micro-Dynamics 188
3.2.1. Transfer functions design and latency management 189
3.2.2. Tactile feedback & haptic perception 189
3.3. Meso-Dynamics 190
3.3.1. Interaction bandwidth and vocabulary 190
3.3.2. Spatial and temporal continuity in interaction 190
3.3.3. Expressive tools for prototyping, studying, and programming interaction 190
3.4. Macro-Dynamics 191
4. Application Domains **192**
5. Highlights of the Year **192**
5.1.1. Personnel 192
5.1.2. Publications 192
5.1.3. Awards 192
6. New Software and Platforms **192**
6.1. ParkEvolution 192
6.2. liblag 193
7. New Results **193**
7.1. Introduction 193
7.2. Micro-dynamics 194
7.2.1. Latency & Transfer functions 194
7.2.2. Understanding touch interaction 195
7.3. Meso-dynamics 195
7.3.1. Improving interaction bandwidth and expressiveness 195
7.3.2. Interacting with specific setups (Large-Displays, Virtual & Augmented Reality) 196
7.3.3. Tools for prototyping and programming interaction 197
7.4. Macro-dynamics 198
7.5. Interaction Machine 198
8. Partnerships and Cooperations **199**
8.1. Regional Initiatives 199
8.2. National Initiatives 199
8.2.1. ANR 199
8.2.1.1. TurboTouch (PRC, 2014-2019) 199
8.2.1.2. Causality (JCJC, 2019-2023) 199
8.2.2. Inria Project Labs 200
8.2.2.1. BCI-LIFT (2015-2019) 200
8.2.2.2. AVATAR (2018-2022) 200

8.2.3. Others 200
8.3. International Initiatives 201
8.4. International Research Visitors 201
8.4.1. Visits of International Scientists 201
8.4.2. Visits to International Teams 201
9. Dissemination **201**
9.1. Promoting Scientific Activities 201
9.1.1. Scientific Events Organisation 201
9.1.2. Scientific Events Selection 201
9.1.2.1. Member of the Conference Program Committees 201
9.1.2.2. Reviewer 201
9.1.3. Journal 202
9.1.3.1. Member of the Editorial Boards 202
9.1.3.2. Reviewer - Reviewing Activities 202
9.1.4. Invited Talks 202
9.1.5. Leadership within the Scientific Community 202
9.1.6. Scientific Expertise 202
9.1.7. Research Administration 202
9.1.7.1. For Inria 202
9.1.7.2. For Inria Lille – Nord Europe 202
9.1.7.3. For the CRIStAL lab of Univ. Lille & CNRS 203
9.1.7.4. For the Université de Lille 203
9.1.8. Hiring committees 203
9.2. Teaching - Supervision - Juries 203
9.2.1. Teaching 203
9.2.2. Supervision 203
9.2.3. Juries 204
9.2.4. Mid-term evaluation committees 204
9.3. Popularization 204
9.3.1. Internal or external Inria responsibilities 204
9.3.2. Articles and contents 204
9.3.3. Education 204
9.3.4. Internal action 204
10. Bibliography **204**

Team LOKI

Creation of the Team: 2018 January 01

Keywords:

Computer Science and Digital Science:

A5.1. - Human-Computer Interaction
A5.1.1. - Engineering of interactive systems
A5.1.2. - Evaluation of interactive systems
A5.1.3. - Haptic interfaces
A5.1.4. - Brain-computer interfaces, physiological computing
A5.1.5. - Body-based interfaces
A5.1.8. - 3D User Interfaces
A5.1.9. - User and perceptual studies
A5.2. - Data visualization
A5.6.1. - Virtual reality
A5.6.4. - Multisensory feedback and interfaces
A5.7.2. - Music

Other Research Topics and Application Domains:

B2.2.6. - Neurodegenerative diseases
B2.8. - Sports, performance, motor skills
B6.1.1. - Software engineering
B9.2.1. - Music, sound
B9.5.1. - Computer science
B9.5.6. - Data science
B9.6.10. - Digital humanities
B9.8. - Reproducibility

1. Team, Visitors, External Collaborators

Research Scientists

Stéphane Huot [Team leader, Inria, Senior Researcher, HDR]
Sylvain Malacria [Inria, Researcher]
Mathieu Nancel [Inria, Researcher]
Marcelo Wanderley [Inria, International Chair & professor at McGill University]

Faculty Members

Géry Casiez [Université de Lille, Professor, HDR]
Thomas Pietrzak [Université de Lille, Associate Professor]

Technical Staff

Damien Masson [Inria, from Oct 2018]

PhD Students

Axel Antoine [Université de Lille]
Marc Baloup [Inria, from Oct 2018]
Amira Chalbi [Inria, until Mar 2018]

Nicole Pong [Inria]
Thibault Raffaillac [Inria & Université de Lille]

Post-Doctoral Fellow

Raiza Sarkis Hanada [Inria, from Dec 2018]

Visiting Scientist

Edward Lank [Inria & Université de Lille, Professor at University of Waterloo, until Aug 2018]

Administrative Assistants

Julie Jonas [Inria, from Jun 2018]
Karine Lewandowski [Inria, until Jun 2018]

2. Overall Objectives

2.1. Introduction

Human-Computer Interaction (HCI) is a constantly moving field [38]. 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 [42]. Different problems and opportunities thus regularly appear that require to be addressed from both the user and the machine perspective, in order to understand and account for the tight coupling between human factors and interactive technologies. Our vision is then to link together these two essential elements: *Knowledge & Technology for Interaction*.

2.2. Knowledge for Interaction

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* we would purposefully use to carry out particular tasks and that would empower people by supporting intelligent use [35]. Others at the same time were seeing computers differently, as *partners*, intelligent entities to whom we would delegate tasks. These two visions still constitute the roots of today's predominant HCI paradigms, *use* and *delegation*. In the delegation approach, a lot of effort has been made to support oral, written and non-verbal forms of human-computer communication, and to analyze and predict human behavior. But the inconsistency and ambiguity of human beings, and the variety and complexity of contexts, make these tasks very difficult [46] and the machine is thus the center of interest.

2.2.1. Computers as tools

Our focus is not in what machines can understand or do by themselves, but in what people can do with them. We do not reject the delegation paradigm but clearly favor the one of tool use, aiming for systems that support intelligent use rather than intelligent systems. And as the frontier between the two is getting thinner, **one of our goals is to better understand what it takes for an interactive system to be perceived as a tool or a partner, and how the two paradigms can be combined for the best benefit of the user.**

2.2.2. Empowering tools

The ability provided by interactive tools to create and control complex transformations in real-time can support intellectual and creative processes in unusual but powerful ways. But mastering powerful tools is not simple and immediate, it requires learning and practice. **Our research in HCI should not just focus on novice or highly proficient users but should also care about intermediate ones willing to devote time and effort to develop new skills, whether for work or leisure.**

2.2.3. Transparent tools

Technology is most empowering when it is transparent: invisible in effect, it does not get into your way but lets you focus on the task. HEIDEGGER characterized this unobtruded relation to things with the term *zuhanden* (*ready-to-hand*). Transparency of interaction is not best achieved with tools mimicking human capabilities, but with those taking full advantage of them given the context and task. For instance, the transparency of driving a car "*is not achieved by having a car communicate like a person, but by providing the right coupling between the driver and action in the relevant domain (motion down the road)*" [49]. Our actions towards the digital world need to be digitized and we must receive proper feedback in return. But input and output technologies pose somewhat inevitable constraints while the number, diversity, and dynamicity of digital objects call 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: Do they leverage our perceptual and control skills? Do they support the right level of coupling for transparent use? Can we improve them or design more suitable ones?**

2.3. Technology for Interaction

Studying the *interactive phenomena* described above is one of the pillars of HCI research, in order to understand, model and ultimately improve them. Yet, we have to make those phenomena happen, to make them possible and reproducible, whether it be for further research or for their diffusion [37]. However, because of the high viscosity and the lack of openness of actual systems, this requires considerable efforts in designing, engineering, implementing and hacking hardware and software interactive artifacts. This is what we call "*The Iceberg of HCI Research*", of which the hidden part supports the design and study of new artifacts, but also informs their creation process.

2.3.1. "Designeering Interaction"

Both parts of this iceberg are strongly influencing each other: The design of interaction techniques informs on the capabilities and limitations of the platform and the software being used, giving insights into what could be done to improve them. On the other hand, new architectures and software tools open the way to new designs, by giving the necessary bricks to build with [39]. These bricks define the adjacent possible of interactive technology, the set of what could be designed by assembling the parts in new ways. Exploring ideas that lie outside of the adjacent possible require the necessary technological evolutions to be addressed first. This is a slow and gradual but uncertain process, which helps to explore and fill a number of gaps in our research field but can also lead to deadlocks. **We want to better understand and master this process –i. e., analyzing the adjacent possible of HCI technology and methods– and introduce tools to support and extend it. This could help to make technology better suited to the exploration of fundamentals of interaction and to their integration into real systems, a way to ultimately improve interactive systems to be empowering tools.**

2.3.2. Computers vs Interactive Systems

In fact, today's interactive systems –e. g., desktop computers, mobile devices– share very similar layered architectures inherited from the first personal computers of the 1970s. This abstraction of resources provides developers with standard components (UI widgets) and high-level input events (mouse and keyboard) that obviously ease the development of common user interfaces for predictable and well-defined tasks and users' behaviors. But it does not favor the implementation of non standard interaction techniques that could be better adapted to more particular contexts, to expressive and creative uses. It often requires to go deeper into the system layers and to hack them until getting access to the required functionalities and/or data, which implies switching between programming paradigms and/or languages.

And these limitations are even more pervading as interactive systems have changed deeply in the last 20 years. They are no longer limited to a simple desktop or laptop computer with a display, a keyboard and a mouse. They are becoming more and more distributed and pervasive (e. g., mobile devices, Internet of Things). They are changing dynamically with recombinations of hardware and software (e. g., transition between multiple devices, modular interactive platforms for collaborative use). Systems are moving "out of the box" with Augmented Reality, and users are going " inside of the box" with Virtual Reality. This is obviously raising new challenges in terms of human factors, usability and design, but it also deeply questions actual architectures.

2.3.3. The Interaction Machine

We believe that promoting digital devices to **empowering tools** requires **better fundamental knowledge about interaction phenomena** AND to **revisit the architecture of interactive systems** in order to support this knowledge. By following a comprehensive systems approach –encompassing human factors, hardware elements and all software layers above– we want to define the founding principles of an *Interaction Machine*:

- a set of hardware and software requirements with associated specifications for interactive systems to be tailored to interaction by leveraging human skills;
- one or several implementations to demonstrate and validate the concept and the specifications in multiple contexts;
- guidelines and tools for designing and implementing interactive systems, based on these specifications and implementations.

To reach this goal, we will adopt an opportunistic and iterative strategy guided by the *designeering* approach, where the engineering part will be fueled by the interaction design and study part. We will address several fundamental problems of interaction related to our vision of "empowering tools", which, in combination with state-of-the-art solutions, will instruct us on the requirements for the solutions to be supported in an interactive system. This consists in reifying the concept of the Interaction Machine in multiple contexts and for multiple problems, before to converge towards a more unified definition of "what is an interactive system", the ultimate Interaction Machine, which makes the main scientific and engineering challenge of our project.

3. Research Program

3.1. Introduction

Interaction is by nature a dynamic phenomenon that takes place between interactive systems and their users. Redesigning interactive systems to better account for interaction requires fine understanding of these dynamics from the user side so as to better handle them from the system side. In fact, layers of actual interactive systems abstract hardware and system resources from a system and programing perspective. Following our Interaction Machine concept, we are reconsidering these architectures from the user perspective, through different *levels of dynamics of interaction* (see Figure 1).

Considering phenomena that occur at each of these levels as well as their relationships will help us to acquire the necessary knowledge (Empowering Tools) and technological bricks (Interaction Machine) to reconcile the way interactive systems are designed and engineered with human abilities. Although our strategy is to investigate issues and address challenges for all of the three levels, our immediate priority is to focus on micro-dynamics since it concerns very fundamental knowledge about interaction and relates to very low-level parts of interactive systems, which is likely to influence our future research and developments at the other levels.

3.2. Micro-Dynamics

Micro-dynamics involve low-level phenomena and human abilities which are related to short time/instantness and to perception-action coupling in interaction, when the user has almost no control or consciousness of the action once it has been started. From a system perspective, it has implications mostly on input and output (I/O) management.

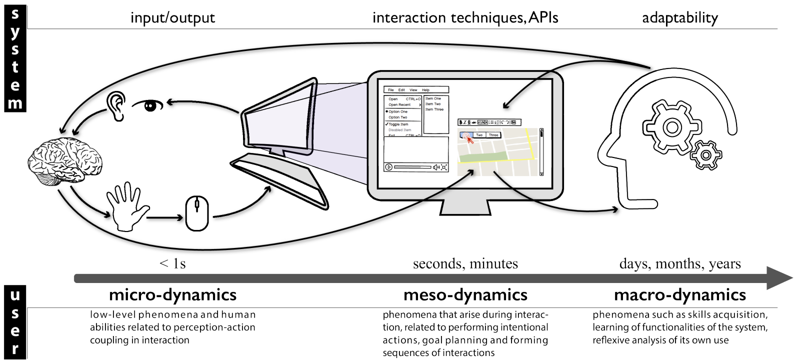

Figure 1. Levels of dynamics of interaction.

3.2.1. Transfer functions design and latency management

We have developed a recognized expertise in the characterization and the design of *transfer functions* [34], [45], i. e., the algorithmic transformations of raw user input for system use. Ideally, transfer functions should match the interaction context. Yet the question of how to maximize one or more criteria in a given context remains an open one, and on-demand adaptation is difficult because transfer functions are usually implemented at the lowest possible level to avoid latency. Latency has indeed long been known as a determinant of human performance in interactive systems [41] and recently regained attention with touch interactions [40]. These two problems require cross examination to improve performance with interactive systems: Latency can be a confounding factor when evaluating the effectiveness of transfer functions, and transfer functions can also include algorithms to compensate for latency.

We have recently proposed new cheap but robust methods for the measurement of end-to-end latency [2] and are currently working on compensation methods and the evaluation of their perceived side effects. Our goal is then to automatically adapt the transfer function to individual users and contexts of use while reducing latency in order to support stable and appropriate control. To achieve this, we will 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. From an engineering perspective, this knowledge on low-level human factors will help us to rethink and redesign the I/O loop of interactive systems in order to better account for them and achieve more adapted and adaptable perception-action coupling.

3.2.2. Tactile feedback & haptic perception

We are also concerned with the physicality of human-computer interaction, with a focus on haptic perception and related technologies. For instance, when interacting with virtual objects such as software buttons on a touch surface, the user cannot feel the click sensation like with physical buttons. The tight coupling between how we perceive and how we manipulate objects is then essentially broken although this is instrumental for efficient direct manipulation. We have addressed this issue in multiple contexts by designing, implementing and evaluating novel applications of tactile feedback [5].

In comparison with many other modalities, one difficulty with tactile feedback is its diversity. It groups sensations of forces, vibrations, friction or deformation. Although this is a richness, it also raises usability and technological challenges since each kind of haptic stimulation requires different kinds of actuators with their own parameters and thresholds. And results from one are hardly applicable to others. On a “knowledge” point of view, we want to better understand and empirically classify haptic variables and the kind of information they can represent (continuous, ordinal, nominal), their resolution, and their applicability to various contexts. From the “technology” perspective, we want to develop tools to inform and ease the design of haptic interactions taking best advantage of the different technologies in a consistent and transparent way.

3.3. Meso-Dynamics

Meso-dynamics relate to phenomena that arise during interaction, on a longer but still short time-scale. For users, it is related to performing intentional actions, to goal planning and tools selection, and to forming sequences of interactions based on a known set of rules or instructions. From the system perspective, it relates to how possible actions are exposed to the user and how they have to be executed (i. e., interaction techniques). It also has implication on the tools for designing and implementing those techniques (programming languages and APIs).

3.3.1. Interaction bandwidth and vocabulary

Interactive systems and their applications have an always increasing number of available features and commands due to e. g., the large amount of data to manipulate, increasing power and number of functionalities, multiple contexts of use.

On the input side, we want to augment the *interaction bandwidth* between the user and the system in order to cope with this increasing complexity. In fact, most input devices capture only a few of the movements and actions the human body is capable of. Our arms and hands for instance have many degrees of freedom that are not fully exploited in common interfaces. We have recently designed new technologies to improve expressibility such as a bendable digitizer pen [36], or reliable technology for studying the benefits of finger identification on multi-touch interfaces [4].

On the output side, we want to expand users’ *interaction vocabulary*. All of the features and commands of a system can not be displayed on screen at the same time and lots of *advanced* features are by default hidden to the users (e. g., hotkeys) or buried in deep hierarchies of command-triggering systems (e. g., menus). As a result, users tend to use only a subset of all the tools the system actually offers [44]. We will study how to help them to broaden their knowledge of available functions.

Through this “opportunistic” exploration of alternative and more expressive input methods and interaction techniques, we will particularly focus on the necessary technological requirements to integrate them into interactive systems, in relation with our redesign of the I/O stack at the micro-dynamics level.

3.3.2. Spatial and temporal continuity in interaction

At a higher-level, we will investigate how such more expressive techniques affect users’ strategies when performing sequences of elementary actions and tasks. More generally, we will explore the “*continuity*” in interaction. Interactive systems have moved from one computer to multiple connected interactive devices (computer, tablets, phones, watches, etc.) that could also be augmented through a Mixed-Reality paradigm. This distribution of interaction raises new challenges from both the usability and engineering perspectives that we obviously have to consider in our main objective of revisiting interactive systems [43]. It involves the simultaneous usage of multiple devices and also the changes in the role of devices according to the location, the time, the task, contexts of use: A tablet device can be used as the main device while traveling, and it becomes an input device or a secondary monitor for continuing the same task once in the office; A smart-watch can be used as a standalone device to send messages, but also as a remote controller for a wall-sized display. One challenge is then to design interaction techniques that support seamless and smooth transitions during these spatial and temporal changes of the system in order to maintain the continuity of uses and tasks, and how to integrate these principles in future interactive systems.

3.3.3. Expressive tools for prototyping, studying, and programming interaction

Actual systems suffers from issues that keep constraining and influencing how interaction is thought, designed, and implemented. Addressing the challenges we presented in this section and making the solutions possible require extended expressiveness, and researchers and designers must either wait for the proper toolkits to appear, or “hack” existing interaction frameworks, often bypassing existing mechanisms. For instance, numerous usability problems in existing interfaces are stemming from a common cause: the lack, or untimely discarding, of relevant information about how events are propagated and changes come to occur in interactive environments. On top of our redesign of the I/O loop of interactive systems, we will investigate how to facilitate access to that information and also promote a more grounded and expressive way to describe and exploit input-to-output chains of events at every system level. We want to provide finer granularity and better-described connections between the *causes* of changes (e.g. input events and system triggers), their *context* (e.g. system and application states), their *consequences* (e.g. interface and data updates), and their *timing* [8]. More generally, a central theme of our Interaction Machine vision is to promote interaction as a first-class object of the system [33], and we will study alternative and better adapted technologies for designing and programming interaction, such as we did recently to ease the prototyping of Digital Musical Instruments [1] or the programming of animations in graphical interfaces [10]. Ultimately, we want to propose a unified model of hardware and software scaffolding for interaction that will contribute to the design of our Interaction Machine.

3.4. Macro-Dynamics

Macro-dynamics concern longer-term phenomena such as skills acquisition, learning of functionalities of the system, reflexive analysis of its own use (e. g., when the user has to face novel or unexpected situations which require high-level of knowledge of the system and its functioning). From the system perspective, it implies to better support cross-application and cross-platform mechanisms so as to favor skill transfer. It also requires to improve the instrumentation and high-level logging capabilities to favor reflexive use, as well as flexibility and adaptability for users to be able to finely tune and shape their tools.

We want to move away from the usual binary distinction between “novices” and “experts” [3] and explore means to promote and assist digital skill acquisition in a more progressive fashion. Indeed, users have a permanent need to adapt their skills to the constant and rapid evolution of the tasks and activities they carry on a computer system, but also the changes in the software tools they use [47]. And software strikingly lacks powerful means of acquiring and developing these skills [3], forcing users to mostly rely on outside support (e. g., being guided by a knowledgeable person, following online tutorials of varying quality). As a result, users tend to rely on a surprisingly limited interaction vocabulary or *make-do* with sub-optimal routines and tools [48]. Ultimately, the user should be able to master the interactive system to form durable and stabilized practices that would eventually become *automatic* and reduce the mental and physical efforts, making their interaction *transparent*.

In our previous work, we identified the fundamental factors influencing expertise development in graphical user interfaces and created a conceptual framework that characterize users’ performance improvement with UIs [7], [3]. We designed and evaluated new command selection and learning methods to leverages user’s digital skill development with user interfaces, on both desktop [6] and touch-based computers.

We are now interested in broader means to support the analytic use of computing tools:

- *to foster understanding of interactive systems.* As the digital world makes the shift to more and more complex systems driven by machine learning algorithms, we increasingly loose our comprehension of what processes yielded the system to respond in one way rather than another. We will study how novel interactive visualizations can help reveal and expose the “intelligence” behind, in ways that people better master their complexity.
- *to foster reflexion on interaction.* We will study how we can foster users’ reflexion on their own interaction in order to encourage them to acquire novel digital skills. We will build real-time and off-line software for monitoring how user’s ongoing activity is conducted at an application and system level. We will develop augmented feedbacks and interactive history visualization tools that will offer contextual visualizations to help users to better understand their activity, compare their actions to that of others, and discover possible improvement.
- *to optimize skill-transfer and tool re-appropriation.* The rapid evolution of new technologies has drastically increased the frequency at which systems are updated, often requiring to relearn everything from scratch. We will explore how we can minimize the cost of having to appropriate an interactive tool by helping users to capitalize on their existing skills when appropriating a new interactive

system.

We plan to explore these questions as well as the use of such aids in several contexts such as web-based, mobile or BCI-based applications. Although, a core aspect of this work will be to design systems and interaction techniques that will be as little platform-specific as possible, in order to better support skill-transfer. Following our Interaction Machine vision, this will lead us to rethink how interactive systems have to be engineered so that they can offer better instrumentation, higher adaptability, and fewer separation between applications and tasks in order to support reuse and skills transfer.

4. Application Domains

4.1. Application Domains

Loki works on fundamental and technological aspects of Human-Computer Interaction that can be applied to diverse application domains.
Our 2018 research concerned desktop, touch-based, haptics, and BCI interfaces with notable applications to medicine (analysis of fine motor control for patients with Parkinson disease), digital humanities (interpretation of handwritten historical documents), as well as creativity support tools (production of illustrations, design of Digital Musical Instruments).

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Personnel

Géry Casiez has been appointed junior member of the Institut Universitaire de France.

Géry Casiez has been appointed at the rank of Adjunct Professor by the University of Waterloo, Canada (2018-2020).

5.1.2. Publications

Loki presented 6 papers at ACM CHI and 1 paper at ACM UIST, the most prestigious conferences in our field.

5.1.3. Awards

"Honorable mention" (top 4% of the 2500+ submissions) from the ACM CHI conference to the paper "Storyboard-Based Empirical Modelling of Touch Interface Performance", from A. Goguey, G. Casiez, A. Cockburn, & C. Gutwin .
BEST PAPERS AWARDS :
[19]
A. GOGUEY, G. CASIEZ, A. COCKBURN, C. GUTWIN.*Storyboard-Based Empirical Modeling of Touch Interface Performance*, in "Adjunct Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018), Demonstration", Montreal, Canada, April 2018 [*DOI :* 10.1145/3170427.3186479], https://hal.inria.fr/hal-01736699

6. New Software and Platforms

6.1. ParkEvolution

Longitudinal analysis of fine motor movement in an ecological context for patients with Parkinson disease
KEYWORD: Parkinson disease

FUNCTIONAL DESCRIPTION: The originality of this application relies on the acquisition of data in an ecological context. Thanks to this application that patients download on their personal computer, the data corresponding to cursor displacement on screen and raw input from pointing devices are collected, encrypted and sent to a server. The analysis of this data allows to compute a motor score according to the parameters of movement, in order to identify alterations in fine motor control. We ensure a realistic score based on the important quantity of data collected. This software is written in C++ and runs on Windows. It uses the libpointing library to access raw data from pointing devices.

RELEASE FUNCTIONAL DESCRIPTION: Fixed bugs and developed new features.
NEWS OF THE YEAR: Release of the 1.3.0 version fixing a number of bugs and introducing a number of new small features. An APP request is close of getting approved.

- Participants: Géry Casiez and Laure Fernandez
- Partners: Aix-Marseille Université - CNRS Laboratoire de Psychologie Cognitive - UMR 7290 - Team 'Perception and attention' - Institut de Neurosciences de la Timone
- Contact: Géry Casiez
- URL: http://parkevolution.org/

6.2. liblag

Library implementing latency compensation techniques for interactive systems
KEYWORDS: Interaction - Latency
FUNCTIONAL DESCRIPTION: The library comprises the management of a set of multitouch input devices, the implementation of latency compensation techniques from the state-of-the-art and new latency compensation techniques developed in the project, and a system to handle artificial latency.

The library is developed in C++ using the Qt framework to allow compiling the same code on a wide range of devices and platforms.
NEWS OF THE YEAR: Finished the work on the latency compensation algorithm "TurboTouch predictor" and development of the demonstrator presented at Euratechnologies. Development of an on-line interactive demo available at http://ns.inria.fr/loki/TTp/

- Contact: Géry Casiez
- Publications: Dispositif à affichage prédictif - Next-Point Prediction for Direct Touch Using Finite-Time Derivative Estimation
- URL: http://mjolnir.lille.inria.fr/turbotouch/

7. New Results

7.1. Introduction

According to our research program, in the next two to five years, we will study dynamics of interaction along three levels depending on interaction time scale and related user's perception and behavior: *Micro-dynamics*, *Meso-dynamics*, and *Macro-dynamics*. Considering phenomena that occur at each of these levels as well as their relationships will help us to acquire the necessary knowledge (Empowering Tools) and technological bricks (Interaction Machine) to reconcile the way interactive systems are designed and engineered with human abilities. Although our strategy is to investigate issues and address challenges for all of the three levels of dynamics, our immediate priority is to focus on micro-dynamics since it concerns very fundamental knowledge about interaction and relates to very low-level parts of interactive systems, which is likely to influence our future research and developments at other levels.

7.2. Micro-dynamics

Participants: Axel Antoine, Géry Casiez [correspondent], Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak.

7.2.1. *Latency & Transfer functions*

End-to-end latency in interactive systems is detrimental to performance and usability, and comes from a combination of hardware and software delays. While these delays are steadily addressed by hardware and software improvements, it is at a decelerating pace. In parallel, short-term input prediction has recently shown promising results to compensate for latency, in both research and industry.

in the context of the collaborative Turbotouch project, we introduced a new prediction algorithm for direct touch devices based on (i) a state-of-the-art finite-time derivative estimator, (ii) a smoothing mechanism based on input speed, and (iii) a post-filtering of the prediction in two steps (see Figure 2 left). Using both a preexisting dataset of touch input as benchmark, and subjective data from a new user study, we showed that this new predictor outperforms those currently available in the literature and industry, based on metrics that model user-defined negative side-effects caused by input prediction. In particular, our predictor can predict up to 2 or 3 times further than existing techniques with minimal negative side-effects [23].

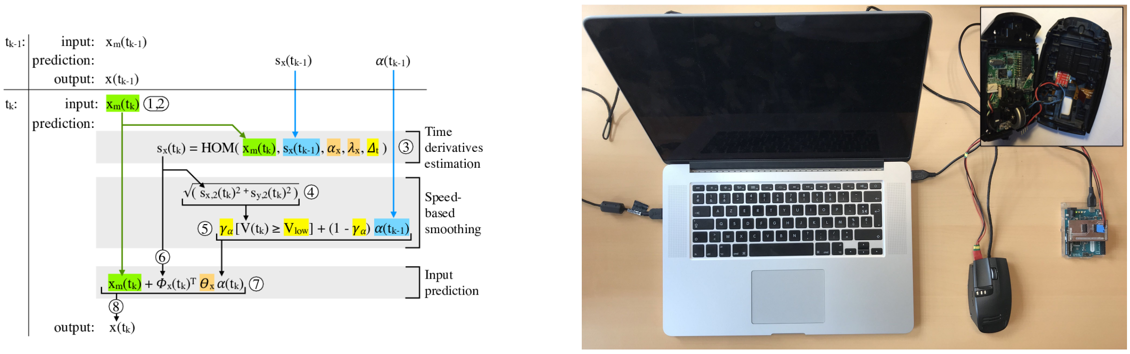

Figure 2. (left) General description of our real-time input prediction method, with step numbers. Input in green, previously computed variables in blue, general parameters in yellow, optimized parameters in orange. (right) Our hybrid setup for input prediction comprises a Logitech G9 Laser Mouse connected via USB to the host computer with the MPU-9250 chip embedded inside, which is itself connected to an Arduino board.

We also proposed a hybrid hardware and software input prediction technique specifically designed for partially compensating end-to-end latency in indirect pointing (see Figure 2 right). We combined a computer mouse with a high frequency accelerometer to predict the future location of the pointer using Euler based equations. Our prediction method results in more accurate prediction than previously introduced prediction algorithms for direct touch. A controlled experiment also revealed that it can improve target acquisition time in pointing tasks [15], [28].

Finally, on the topic of transfer functions we performed some preliminary analysis of the kinematics of a pointing task with varying linear velocity based transfer functions to assess how we use vision and haptics to plan and control our movement [25].

7.2.2. Understanding touch interaction

Atomic interactions in touch interfaces, like tap, drag, and flick, are well understood in terms of interaction design, but less is known about their physical performance characteristics. We conducted a study to gather baseline data about finger pitch and roll orientation during atomic touch input actions [21]. Our results showed differences in orientation and range for different fingers, hands, and actions: for a given hand, the little, ring and middle fingers are used in a similar manner, whereas the thumb uses different range of orientations. Additional analyses about how changing the angle of the tablet affects people's finger orientations suggest that ranges of orientation tighten as the tablet pitch increases. This data provides designers and researchers with better understanding of what kind of interactions are possible in different settings (e. g., using the left or right hand), to design novel interaction techniques that use orientation as input (e. g., using finger tilt as an implicit mode), and to anticipate the feasibility of new sensing techniques (e. g., using fingerprints for identifying specific finger touches).

7.3. Meso-dynamics

Participants: Marc Baloup, Géry Casiez, Stéphane Huot, Edward Lank, Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak [correspondent], Thibault Raffaillac, Marcelo Wanderley.

7.3.1. Improving interaction bandwidth and expressiveness

Despite the ubiquity of touch-based input and the availability of increasingly computationally powerful touchscreen devices, there has been comparatively little work on enhancing basic canonical gestures such as swipe-to-pan and pinch-to-zoom. We introduced transient pan and zoom, i. e., pan and zoom manipulation gestures that temporarily alter the view and can be rapidly undone [16]. Leveraging typical touchscreen support for additional contact points, we designed our transient gestures so that they co-exist with traditional pan and zoom interaction. In addition to reducing repetition in multi-level navigation, our transient pan-and-zoom also facilitates rapid movement between document states.

Image editing software feature various pixel selection tools based on geometrical (rectangle, ellipses, polygons) or semantical (magic wand, selection brushes) data from the image. They are efficient in many situations, but are limited when selecting bitmap representations of handwritten text for e. g., interpreting scanned historical documents that cannot be reliably analyzed by automatic OCR methods: strokes are thin, with many overlaps and brightness variations. We have designed a new selection tool dedicated to this purpose [27]: a cursor based brush selection tool with two additional degrees of freedom: brush size and brightness threshold. The brush cursor displays feedforward clues that indicates the user which pixels will be selected upon pressing the mouse button. This brush provides a fine grain control to the user over the selection.

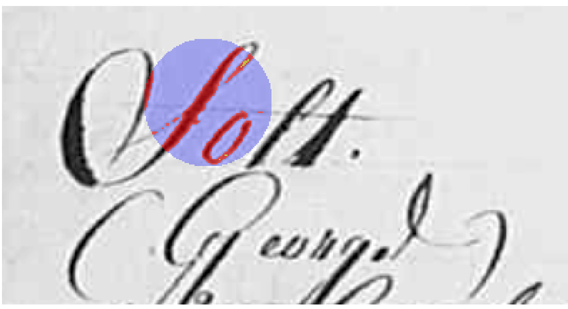

Figure 3. A four-dimensional selection brush for digitized handwritten documents. Red pixels will be selected, blue pixels will not.

7.3.2. Interacting with specific setups (Large-Displays, Virtual & Augmented Reality)

Large displays are becoming commonplace at work, at home, or in public areas. Handheld devices such as smartphones and smartwatches are ubiquitous, but little is known on regarding how these devices could be used to point at remote large displays. We conducted a survey on possession and use of smart devices, as well as a controlled experiment comparing seven distal pointing techniques on phone or watch, one- and two-handed, and using different input channels and mappings [26]. Our results favor using a smartphone as a trackpad, but also explore performance tradeoffs that can inform the choice and design of distal pointing techniques for different contexts of use.

In virtual reality environments, raycasting is the most common target pointing technique. However, performance on small and distant targets is impacted by the accuracy of the pointing device and the user's motor skills. Existing pointing facilitation techniques are currently only applied in the context of a virtual hand, i. e., for targets within reach. We studied how a user-controlled cursor could be added on the ray in order to enable target proximity-based pointing techniques –such as the Bubble Cursor– to be used for targets that are out of reach [17]. We conducted a study comparing several visual feedbacks for this technique (see Figure 4). Our results showed that simply highlighting the nearest target reduces the selection time by 14.8% and the error rate by 82.6% compared to standard Raycasting. For small targets, the selection time is reduced by 25.7% and the error rate by 90.8%.

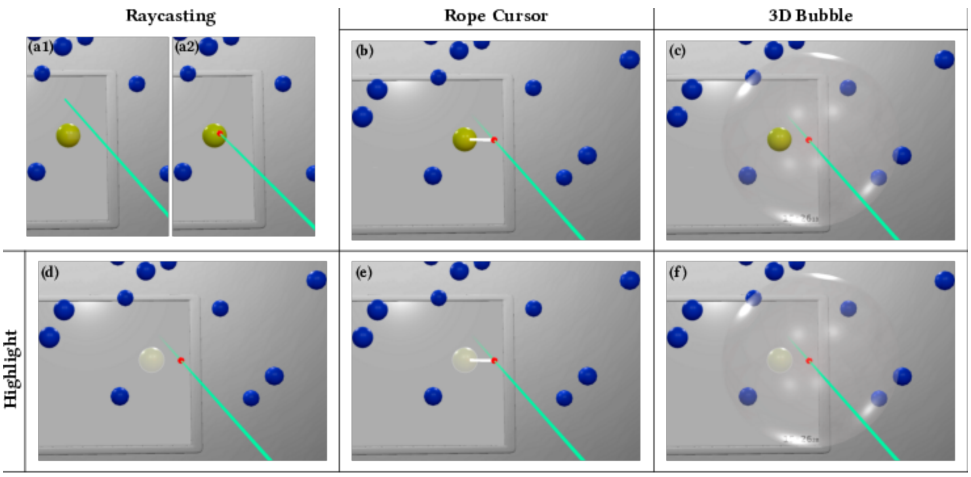

Figure 4. Visual feedback for RayCursor: (a1,a2) classical Raycasting; (b) Rope Cursor: a stroke between the closest target and the cursor; (c) 3D Bubble: a bubble centered on the cursor which contains the nearest target; (d) Highlighting on the nearest target; (e,f), highlight + rope and 3D Bubble.

Brain-Computer Interfaces (BCIs) enable users to interact with computers without any dedicated movement, bringing new hands-free interaction paradigms that could be beneficial in an Augmented Reality (AR) setup. We first tested the feasibility of using BCI in AR settings based on Optical See-Through Head-Mounted Displays (OST-HMDs) [12]. Experimental results showed that a BCI and an OST-HMD equipment (EEG headset and Hololens in our case) are well compatible and that small movements of the head can be tolerated when using the BCI. Then, we introduced a design space for command display strategies based on BCI in AR, when exploiting a famous brain pattern called Steady-State Visually Evoked Potential (SSVEP). Our design space relies on five dimensions concerning the visual layout of the BCI menu: orientation, frame-of-reference, anchorage, size and explicitness. We implemented various BCI-based display strategies and tested them within the context of mobile robot control in AR. Our findings were finally integrated within an operational prototype based on a real mobile robot that is controlled in AR using a BCI and a HoloLens headset. Taken together, our results (4 user studies) and our methodology could pave the way to future interaction schemes in Augmented Reality exploiting 3D User Interfaces based on brain activity and BCIs.

More generally, we also contributed to a reflexion on the complexity and scientific challenges associated to virtual and augmented realities [29] and the challenges to make virtual environments more closely related to the real world [30].

7.3.3. Tools for prototyping and programming interaction

Touch interactions are now ubiquitous, but few tools are available to help designers quickly prototype touch interfaces and predict their performance. On one hand, for rapid prototyping, most applications only support visual design. On the other hand, for predictive modeling, tools such as CogTool generate performance predictions but do not represent touch actions natively and do not allow exploration of different usage contexts. To combine the benefits of rapid visual design tools with underlying predictive models, we developed the *Storyboard Empirical Modeling (StEM)* tool [20], [19] for exploring and predicting user performance with touch interfaces (see Figure 5). StEM provides performance models for mainstream touch actions, based on a large corpus of realistic data. We evaluated StEM in an experiment and compared its predictions to empirical times for several scenarios. The study showed that our predictions are accurate (within 7% of empirical values on average), and that StEM correctly predicted differences between alternative designs. Our tool provides new capabilities for exploring and predicting touch performance, even in the early stages of design.

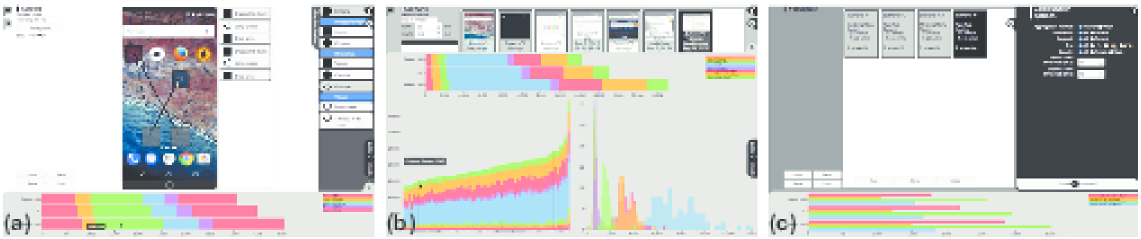

Figure 5. Storyboard Empirical Modeling (StEM): (a) users drag and drop actions onto a timeline to construct an interaction sequence; (b) users can visualize prediction times for a scenario composed of different screens; (c) users can compare scenarios, and filter the predictions according to contextual factors such as screen size or user's expertise.

Following our main objective of revisiting interactive system, we have also proposed two systems for defining and programming interactive behaviors and interactions.

Much progress has been made on interactive behavior development tools for expert programmers. However, less effort has been made in investigating how these tools support creative communities who typically struggle with technical development. This is the case, for instance, of media artists and composers working with interactive environments. To address this problem, we have introduces ZenStates [18], a new specification model for creative interactive environments that combines Hierarchical Finite-States Machines, expressions, off-the-shelf components called Tasks, and a global communication system called the Blackboard. We have implemented our model in a direct manipulation-based software interface and probed ZenStates' expressive power through 90 exploratory scenarios. we have also conducted a user study to investigate the understandability of ZenStates' model. Results support ZenStates viability, its expressiveness, and suggest that ZenStates is easier to understand –in terms of decision time and decision accuracy– compared to popular alternatives such as standard object-oriented programming and a data-flow visual language.

In a more general context, we have introduced a new GUI framework based on the *Entity-Component-System* model (ECS), where interactive elements (Entities) can acquire any data (Components) [24]. Behaviors are managed by continuously running processes (Systems) which select entities by the components they possess. This model facilitates the handling and reuse of behaviors. It allows to define the interaction modalities of an application globally, by formulating them as a set of Systems. We have implemented an experimental toolkit based on this approach, *Polyphony*, in order to demonstrate the use and benefits of this model.

7.4. Macro-dynamics

Participants: Stéphane Huot, Sylvain Malacria [correspondent], Nicole Pong.

One conspicuous feature of the current evolution of interactive devices is the spread of touch-sensitive surfaces. Typically, modern smartphones are equipped with such touch-sensitive surfaces that also support normal force-based input capabilities, which can for instance be used to control the range of a text selection by varying the force applied to the touchscreen (on e. g., iOS devices). However, this interaction mechanism is difficult to discover and many users simply ignore it exists. To overcome this problem, we introduced ForceSelect (see Figure 6, left), a force-based text selection techniques that relies on a simple mode gauge (see Figure 6, right) that does not require additional screen real-estate and help users to discover and master the use of force input in text selection tasks [22]. We conducted two studies that suggest that this mode gauge successfully provides enhanced discoverability of the force-based input and combines support for novices and experts, whereas it was never worse than the standard iOS technique and was also preferred by participants.

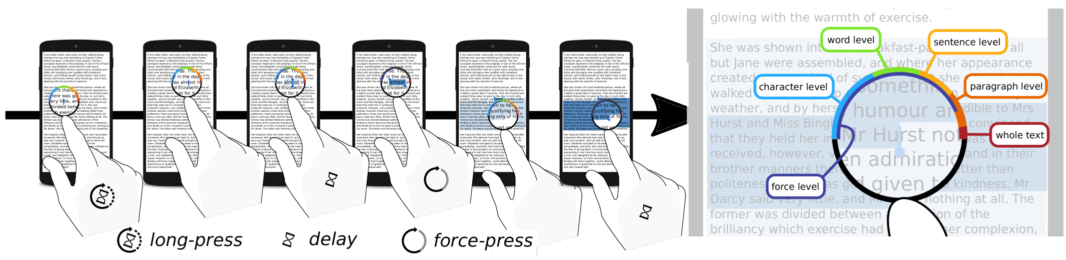

Figure 6. (left) Example of text selection using ForceSelect. The user performs a long-press that displays the callout magnifier. Keeping the force in the character level, the user adjusts its position by moving her finger. She then holds the force in the word level of the "mode gauge", locks the selection and enters the clutch mode. When force-pressing to the whole text level of the "mode gauge", she un-clutches the selection and updates it.; (right) Close-up of the "mode gauge". There are two types of text highlighting in the background: dark highlighting covers between both handles and light highlighting acts as a feedforward of which portion of text will be selected if the user released her finger (here the whole paragraph).

7.5. Interaction Machine

Several of our new results this year contributed to our global objective of building an Interaction Machine, especially at the micro-dynamics level. Our work on prediction algorithms and our hybrid hardware-software latency compensation method highlighted the need for accessing low-level input data and to have flexible input management to be able to reliably predict current finger position and compensate for latency. Our work on the characterization of the dimensions of touch interaction, especially angle of touch, highlighted the need for additional dimensions in input events that are not yet accessible in actual systems. All in all, this confirm our hypothesis that we have to redefine input management and input events propagation in order to better account for human factors in interactive systems, to extend the possibilities for designing more efficient and expressive interaction methods.

At the meso-dynamics level, our work on improving basic interaction methods in non-standard setups (e. g., VR, AR) highlighted the need for more open and flexible system architectures and tools that ease the design and prototyping of alternative interaction techniques based on mixed modalities. The new prototyping and programing tools that we proposed this year (StEM, ZenStates and Polyphony) are our first explorations toward such system-integrated frameworks dedicated to interaction.

8. Partnerships and Cooperations

8.1. Regional Initiatives

8.1.1. GeneaLire (CPER MAuVE, 2018-2019)

Participants: Stéphane Huot, Thomas Pietrzak [correspondent].

Interactive tools for the interpretation of manuscripts

The goal of this project is to design, implement and evaluate interactive tools for helping transcription of scanned handwritten documents. Current solutions focus on automatic recognition, with recent advances thanks to deep learning methods. However these solutions still require a significant learning base that has to be made by hand. Not only this means that part of the work cannot be done automatically, but it also means that this technique is not a solution for small collections of documents. The tools we propose to create will ingeniously take advantage of interactive and automatic techniques. The interactive tools include a text selection techniques [27], as well as advanced annotation techniques that will support collaborative work. This tool will be invaluable for bootstrapping the transcription of large collections, as well as helping transcribing small collections. We will use user-centered design, in order to make sure the tool fits historians and genealogists activities and workflow.

Partners: Inria Saclay's AVIZ team, École Polytechnique de l'Université de Tours, Laboratoire de Démographie et d'Histoire Sociale at l'École des hautes études en sciences sociales, and Geneanet.

Related publication: [27]

8.2. National Initiatives

8.2.1. ANR

8.2.1.1. TurboTouch (PRC, 2014-2019)

Participants: Géry Casiez [correspondent], Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak.

High-performance touch interactions

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 in 2018: [28], [15], [23], [25]

8.2.1.2. Causality (JCJC, 2019-2023)

Participant: Mathieu Nancel [correspondent].

Integrating Temporality and Causality to the Design of Interactive Systems

The project addresses a fundamental limitation in the way interfaces and interactions are designed and even thought about today, an issue we call *procedural information loss*: once a task has been completed by a computer, significant information that was used or produced while processing it is rendered inaccessible regardless of the multiple other purposes it could serve. It hampers the identification and solving of identifiable usability issues, as well as the development of new and beneficial interaction paradigms. We will explore, develop, and promote finer granularity and better-described connections between the causes of those changes, their context, their consequences, and their timing. We will apply it to facilitate the real-time detection, disambiguation, and solving of frequent timing issues related to human reaction time and system latency; to provide broader access to all levels of input data, therefore reducing the need to "hack" existing frameworks to implement novel interactive systems; and to greatly increase the scope and expressiveness of command histories, allowing better error recovery but also extended editing capabilities such as reuse and sharing of previous actions.

Web site: http://loki.lille.inria.fr/causality/

8.2.2. Inria Project Labs

8.2.2.1. BCI-LIFT (2015-2019)

Participant: Géry Casiez [correspondent].

Brain Computer Interfaces: Learning, Interaction, Feedback, Training

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 publication in 2018: [12]

8.2.2.2. AVATAR (2018-2022)

Participants: Géry Casiez, Stéphane Huot, Thomas Pietrzak [correspondent].

The next generation of our virtual selves in digital worlds

This project aims at delivering the next generation of virtual selves, or *avatars*, in digital worlds. In particular, we want to push further the limits of perception and interaction through our avatars to obtain avatars that are better embodied and more interactive. Loki's contribution in this project consists in designing novel 3D interaction paradigms for avatar-based interaction and to design new multi-sensory feedbacks to better feel our interactions through our avatars.

Partners: Inria's GRAPHDECO, HYBRID, MIMETIC, MORPHEO & POTIOC teams, Mel Slater (Event Lab, University Barcelona, Spain), Technicolor and Faurecia.

Web site: https://avatar.inria.fr/

Related publication in 2018: [17]

8.2.3. Others

8.2.3.1. ParkEvolution (Carnot Inria - Carnot STAR, 2015-2019)

Participant: Géry Casiez [correspondent].

Longitudinal analysis of fine motor control for patients with Parkinson disease

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/

8.3. International Initiatives

8.3.1. Inria International Partners

8.3.1.1. Informal International Partners

Andy Cockburn, University of Canterbury, Christchurch, NZ [19], [20]

Carl Gutwin, University of Saskatchewan, Saskatoon, CA [19], [20], [21], [22]

Nicolai Marquardt, University College London, London, UK

Antti Oulasvirta, Aalto University, Helsinki, FI [31]

Daniel Vogel, University of Waterloo, Waterloo, CA [21]

8.4. International Research Visitors

8.4.1. Visits of International Scientists

Edward Lank, Professor at the University of Waterloo, has spent two years in our team until Aug. 2018 (funded by Région Hauts-de-France, Université Lille and Inria).

Marcelo Wanderley, Professor at McGill University, who has been awarded an Inria International Chair in our team in 2016, spent 3 months in our group this year (July to September).

8.4.2. Visits to International Teams

8.4.2.1. Research Stays Abroad

Géry Casiez has spent four months in the Human Computer Interaction Lab at the University of Waterloo (September to December).

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Organisation

9.1.1.1. Member of the Organizing Committees

- IHM (AFIHM): Stéphane Huot (Doctoral Consortium co-Chair), Thomas Pietrzak (Posters co-Chair)

9.1.2. Scientific Events Selection

9.1.2.1. Member of the Conference Program Committees

- CHI (ACM): Géry Casiez
- IHM (AFIHM): Thomas Pietrzak

9.1.2.2. Reviewer

- CHI (ACM): Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak
- UIST (ACM): Géry Casiez, Sylvain Malacria, Mathieu Nancel, Thomas Pietrzak
- MobileHCI (ACM): Thomas Pietrzak
- ISS (ACM): Mathieu Nancel
- NordiCHI: Mathieu Nancel, Thomas Pietrzak

- Eurohaptics: Thomas Pietrzak
- IHM (AFIHM): Stéphane Huot, Mathieu Nancel
- NIME: Marcelo Wanderley
- GI: Mathieu Nancel

9.1.3. Journal

9.1.3.1. Member of the Editorial Boards

- Computer Music Journal: Marcelo Wanderley

9.1.3.2. Reviewer - Reviewing Activities

- Transactions on Computer-Human Interaction (ACM): Mathieu Nancel
- Virtual Environments (Frontiers): Géry Casiez
- IEEE Access (IEEE): Géry Casiez
- Behaviour & Information Technology (Taylor & Francis): Mathieu Nancel

9.1.4. Invited Talks

- *Human-Computer Interaction: Back to the future and... forward to the past*, Insights on the FUTURE of Computing conferences, Laboratoire d'Informatique de Grenoble, Grenoble: Stéphane Huot
- *Interaction Homme-Machine : passé composé et futur simple... ou l'inverse*, Collège de France Seminar – lecture of Gérard Berry (chair Algorithms, Machines and languages), Lille: Stéphane Huot
- *The measure and compensation of latency in touch and mouse-based systems*, DGP, Toronto, Canada: Géry Casiez
- *The measure and compensation of latency in touch and mouse-based systems*, Google, Waterloo, Canada: Géry Casiez
- *Motion Capture of Music Performances: Overview of almost 2 decades of research*, RITMO International Motion Capture Workshop, Oslo, Norway: Marcelo Wanderley (Keynote Speaker)

9.1.5. Leadership within the Scientific Community

Association Francophone d'Interaction Homme-Machine (AFIHM):

- Stéphane Huot: member of the scientific council
- Thomas Pietrzak: board member and webmaster until Oct. 2018

9.1.6. Scientific Expertise

- Agence Nationale de la Recherche: Stéphane Huot (mid-term review committee member)
- CN35 AFNOR normalization committee about normalizing the French keyboard, in collaboration with Aalto University and the Max Planck Institute: Mathieu Nancel
- FWO Research Foundation - Flanders: Géry Casiez (reviewer)
- Région Aquitaine: Géry Casiez, Sylvain Malacria (reviewer)

9.1.7. Research Administration

9.1.7.1. For Inria

- International relations working group (COST-GTRI): Stéphane Huot (member)

9.1.7.2. For Inria Lille – Nord Europe

- "Bureau du comité des équipes projets" (BCEP): Stéphane Huot (member)
- Research jobs committee (CER): Sylvain Malacria (member since Jan. 2018)
- Operational legal and ethical risk assessment committee (COERLE): Stéphane Huot (local correspondent)

9.1.7.3. For the CRIStAL lab of Univ. Lille & CNRS

- Direction board: Géry Casiez (deputy director since Sept. 2018)
- Laboratory council: Géry Casiez (member until Aug. 2018)
- Computer Science PhD recruiting committee: Géry Casiez (member)

9.1.7.4. For the Université de Lille

- Coordinator for internships at IUT A: Géry Casiez
- Computer Science Department council: Thomas Pietrzak

9.1.8. Hiring committees

- Inria's eligibility jury for Junior Researcher Positions (CRCN) in Lille: Stéphane Huot (vice-president)
- Université de Lille hiring committee for a Computer Science Assistant Professor position: Géry Casiez (president)
- Université Paris-Saclay hiring committees for Computer Science Assistant Professor positions: Géry Casiez (member), Stéphane Huot (member)

9.2. Teaching - Supervision - Juries

9.2.1. Teaching

DUT Informatique: Géry Casiez (38h), Stéphane Huot (28h), *IHM*, 1st year, IUT A de Lille - Univ. Lille

Licence Informatique: Thomas Pietrzak, *Logique*, 52.5h, L3, Univ. Lille

Licence Informatique: Thomas Pietrzak, *Automates et Langages*, 36h, L3, Univ. Lille

Licence Informatique: Thomas Pietrzak, *Image et Interaction 2D*, 10.5h, L3, Univ. Lille

Licence Sciences pour l'Ingénieur (SPI): Sylvain Malacria, *Introduction à l'Interaction Homme Machine*, 30h, L3, Institut Villebon Georges Charpak

Cursus ingénieur: Sylvain Malacria (10h), *3DETech*, IMT Lille-Douai

Master Informatique: Thomas Pietrzak (18h), *NIHM*, M2, Univ. Lille

Master Informatique: Sylvain Malacria (20h), *NIHM*, M2, Univ. Lille

Master Informatique: Géry Casiez (4h), Thomas Pietrzak (4h), *Projets*, M2, Univ. Lille

Master Informatique: Thomas Pietrzak (34.5h), Sylvain Malacria (34.5), *IHM*, M1, Univ. Lille

Master Informatique: Mathieu Nancel, *Evaluation*, 4h, M2, Univ. Lille

Master: Thomas Pietrzak (10.5h), *3DETech : 3D Digital Entertainment Technologies*, M2, IMT Lille Douai

Doctorat: Géry Casiez, *Expériences contrôlées et analyses statistiques*, École doctorale SPI

Master & Doctorat: Marcelo Wanderley (39h), MUMT-619 Input Devices for Musical Expression

9.2.2. Supervision

PhD: Amira Chalbi, *Understanding and Designing Animations in the User Interfaces*, defended in April 2018 [11], advised by Nicolas Roussel & Fanny Chevalier

PhD in progress: Marc Baloup, *Interaction with avatars in immersive virtual environments*, started Oct. 2018, advised by Géry Casiez & Thomas Pietrzak

PhD in progress: Axel Antoine, *Helping Users with Interactive Strategies*, started Oct. 2017, advised by Géry Casiez & Sylvain Malacria

PhD in progress: Nicole Ke Chen Pong, *Understanding and Improving Users Interactive Vocabulary*, started Oct. 2016, advised by Nicolas Roussel, Sylvain Malacria & Stéphane Huot

PhD in progress: Thibault Raffaillac, *Languages and System Infrastructure for Interaction*, started Oct. 2015, advised by Stéphane Huot

PhD in progress: Hakim Si Mohammed, *Improving Interaction Based on a Brain-Computer Interface*, started Oct. 2016, advised by Anatole Lecuyer, Ferran Argelaguet, Géry Casiez & Nicolas Roussel (in Rennes)

PhD in progress: Jeronimo Barbosa, *Design and Evaluation of Digital Musical Instruments*, McGill University, started in 2013, advised by Marcelo Wanderley & Stéphane Huot (in Montréal)

9.2.3. Juries

Jeff Avery (PhD, Univ. Waterloo, Canada): Géry Casiez, examiner

Antoine Costes (PhD, Univ. Bretagne Loire): Géry Casiez, reviewer

Bruno Fruchard (PhD, Univ. Paris-Saclay): Géry Casiez, reviewer

Julien Gori (PhD, Univ. Paris-Saclay): Stéphane Huot, examiner

Wanyu Liu (PhD, Univ. Paris-Saclay): Géry Casiez, reviewer

Gary Perelman (PhD, Univ. Toulouse III): Géry Casiez, reviewer

Houssem Saidi (PhD, Univ. Toulouse III): Stéphane Huot, reviewer

9.2.4. Mid-term evaluation committees

- Jeff Avery (PhD, Univ. Waterloo, Canada): Géry Casiez
- Delphine Poux (PhD, Univ. Lille): Géry Casiez
- Marc Teyssier (PhD, Univ. Paris-Saclay): Géry Casiez

9.3. Popularization

9.3.1. Internal or external Inria responsibilities

AIRLab selection committee for the funding of art and science projects, Stéphane Huot (representative for Inria Lille – Nord Europe)

9.3.2. Articles and contents

2067 ou la disparition des interfaces humains-machines, interview in an Inriality and Usbek & Rica joint article, Stéphane Huot

Le raccourci clavier: Une meilleure interaction entre l'Homme et la machine ?, in Lille by Inria N^{O}7, Sylvain Malacria

9.3.3. Education

Structuration et fonctionnement de la recherche, training session for academic librarians at MédiaLille, Thomas Pietrzak

9.3.4. Internal action

Chorégraphie d'animation : structurer pour mieux comprendre, talk at Inria Lille – Nord Europe "13:45", Amira Chalbi

Reliability and Perceived Reliability in Ubiquitous Gestural Input, talk at Inria Lille – Nord Europe "30 minutes de sciences", Edward Lank

The TurboTouch latency compensation method, demonstration at the opening ceremony of Bâtiment Place (Inria Lille – Nord Europe), Mathieu Nancel

10. Bibliography

Major publications by the team in recent years

[1] F. CALEGARIO, M. WANDERLEY, S. HUOT, G. CABRAL, G. RAMALHO.*A method and toolkit for digital musical instruments: generating ideas and prototypes*, in "IEEE MultiMedia", January 2017, vol. 24, n^{O} 1, p. 63-71, https://doi.org/10.1109/MMUL.2017.18

[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] 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

[4] A. GOGUEY, D. VOGEL, F. CHEVALIER, T. PIETRZAK, N. ROUSSEL, G. CASIEZ.*Leveraging finger identification to integrate multi-touch command selection and parameter manipulation*, in "International Journal of Human-Computer Studies", March 2017, vol. 99, p. 21-36, http://dx.doi.org/10.1016/j.ijhcs.2016.11.002

[5] 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

[6] 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

[7] S. MALACRIA, J. SCARR, A. COCKBURN, C. GUTWIN, T. GROSSMAN.*Skillometers: reflective widgets that motivate and help users to improve performance*, in "Proceedings of UIST'13", ACM, October 2013, p. 321-330, http://doi.acm.org/10.1145/2501988.2501996

[8] M. NANCEL, A. COCKBURN.*Causality: a conceptual model of interaction history*, in "Proceedings of CHI'14", ACM, April 2014, p. 1777-1786, http://dx.doi.org/10.1145/2556288.2556990

[9] M. NANCEL, D. VOGEL, B. DE ARAÙJO, R. JOTA, G. CASIEZ.*Next-point prediction metrics for perceived spatial errors*, in "Proceedings of UIST'16", ACM, October 2016, p. 271-285, http://dx.doi.org/10.1145/2984511.2984590

[10] T. RAFFAILLAC, S. HUOT, S. DUCASSE.*Turning Function Calls Into Animations*, in "Proceedings of EICS'17", ACM, June 2017, 6, http://doi.acm.org/10.1145/3102113.3102134

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] A. CHALBI.*Understanding and designing animations in the user interfaces*, Université lille1, April 2018, https://hal.archives-ouvertes.fr/tel-01881889

Articles in International Peer-Reviewed Journal

[12] H. SI-MOHAMMED, J. PETIT, C. JEUNET, F. ARGELAGUET-SANZ, F. SPINDLER, A. EVAIN, N. ROUSSEL, G. CASIEZ, A. LÉCUYER.*Towards BCI-based Interfaces for Augmented Reality: Feasibility, Design and Evaluation*, in "IEEE Transactions on Visualization and Computer Graphics", October 2018, p. 1-12 [*DOI :* 10.1109/TVCG.2018.2873737], https://hal.inria.fr/hal-01947344

[13] O. ZINENKO, S. HUOT, C. BASTOUL.*Visual Program Manipulation in the Polyhedral Model*, in "ACM Transactions on Architecture and Code Optimization", March 2018, vol. 15, n^o 1, p. 1 - 25 [*DOI :* 10.1145/3177961], https://hal.inria.fr/hal-01744426

International Conferences with Proceedings

[14] A. ANTOINE.*Understanding and improving the relation between interaction simplicity and expressivity for producing illustrations*, in "IHM 2018 - 30eme conférence francophone sur l'Interaction Homme-Machine", Brest, France, Rencontres Doctorales, AFIHM, October 2018, p. 1-6, https://hal.archives-ouvertes.fr/hal-01900071

[15] A. ANTOINE, S. MALACRIA, G. CASIEZ.*Using High Frequency Accelerometer and Mouse to Compensate for End-to-end Latency in Indirect Interaction*, in "Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018)", Montréal, Canada, April 2018, p. 1-11 [*DOI :* 10.1145/3173574.3174183], https://hal.inria.fr/hal-01714204

[16] J. AVERY, S. MALACRIA, M. NANCEL, G. CASIEZ, E. LANK.*Introducing Transient Gestures to Improve Pan and Zoom on Touch Surfaces*, in "Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018)", Montréal, Canada, April 2018, p. 1-8 [*DOI :* 10.1145/3173574.3173599], https://hal.inria.fr/hal-01714269

[17] M. BALOUP, V. OUDJAIL, T. PIETRZAK, G. CASIEZ.*Pointing Techniques for Distant Targets in Virtual Reality*, in "Proceedings of the AFIHM Conférence Francophone sur l'interaction Homme-Machine (IHM 2018)", Brest, France, Articles Scientifiques, AFIHM, October 2018, 8, https://hal.archives-ouvertes.fr/hal-01899061

[18] J. BARBOSA, M. M. WANDERLEY, S. HUOT.*ZenStates: Easy-to-Understand Yet Expressive Specifications for Creative Interactive Environments*, in "VL/HCC 2018 - IEEE Symposium on Visual Languages and Human-Centric Computing", Lisbon, Portugal, October 2018, https://hal.inria.fr/hal-01888802

[19] *Best Paper*
A. GOGUEY, G. CASIEZ, A. COCKBURN, C. GUTWIN.*Storyboard-Based Empirical Modeling of Touch Interface Performance*, in "Adjunct Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018), Demonstration", Montreal, Canada, April 2018 [*DOI :* 10.1145/3170427.3186479], https://hal.inria.fr/hal-01736699.

[20] A. GOGUEY, G. CASIEZ, A. COCKBURN, C. GUTWIN.*Storyboard-Based Empirical Modelling of Touch Interface Performance*, in "Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018)", Montréal, Canada, ACM, April 2018 [*DOI :* 10.1145/3173574.3174019], https://hal.archives-ouvertes.fr/hal-01714825

[21] A. GOGUEY, G. CASIEZ, D. VOGEL, C. GUTWIN.*Characterizing Finger Pitch and Roll Orientation During Atomic Touch Actions*, in "Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018)", Montréal, Canada, April 2018, p. 1-12 [*DOI :* 10.1145/3173574.3174163], https://hal.archives-ouvertes.fr/hal-01714827

[22] A. GOGUEY, S. MALACRIA, C. GUTWIN.*Improving Discoverability and Expert Performance in Force-Sensitive Text Selection for Touch Devices with Mode Gauges*, in "CHI 2018 - ACM Conference on Human Factors in Computing Systems", Montreal, Canada, April 2018, p. 1-12 [*DOI :* 10.1145/3173574.3174051], https://hal.inria.fr/hal-01714195

[23] M. NANCEL, S. ARANOVSKIY, R. USHIROBIRA, D. EFIMOV, S. POULMANE, N. ROUSSEL, G. CASIEZ.*Next-Point Prediction for Direct Touch Using Finite-Time Derivative Estimation*, in "Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2018)", Berlin, Germany, October 2018 [*DOI :* 10.1145/3242587.3242646], https://hal.inria.fr/hal-01893310

[24] T. RAFFAILLAC, S. HUOT.*Applying the Entity-Component-System Model to Interaction Programming*, in "IHM 2018 - 30eme conférence francophone sur l'Interaction Homme-Machine", Brest, France, Articles Scientifiques, AFIHM, October 2018, p. 42-51, https://hal.archives-ouvertes.fr/hal-01898859

[25] C. SCOTTO DI CESARE, V. H. VU, G. CASIEZ, L. FERNANDEZ.*Sensorimotor control and linear visuohaptic gain*, in "EuroHaptics 2018 - 11th International Conference on Haptics: Science, Technology, and Applications", Pisa, Italy, June 2018, https://hal.archives-ouvertes.fr/hal-01944914

[26] S. SIDDHPURIA, S. MALACRIA, M. NANCEL, E. LANK.*Pointing at a Distance with Everyday Smart Devices*, in "Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018)", Montréal, Canada, April 2018, p. 1-11 [*DOI :* 10.1145/3173574.3173747], https://hal.inria.fr/hal-01714274

Conferences without Proceedings

[27] E. ALAWOE, T. PIETRZAK, S. HUOT.*Outil de sélection de texte manuscrit sur des documents numérisés*, in "SIFED 2018 - Symposium International Francophone sur l'Ecrit et le Document", Tours, France, May 2018, https://hal.archives-ouvertes.fr/hal-01899373

[28] A. ANTOINE, S. MALACRIA, G. CASIEZ.*TurboMouse: End-to-end Latency Compensation in Indirect Interaction*, in "Adjunct Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2018), Demonstration", Montreal, Canada, April 2018 [*DOI :* 10.1145/3170427.3186542], https://hal.inria.fr/hal-01726763

Scientific Books (or Scientific Book chapters)

[29] F. ARGELAGUET-SANZ, B. ARNALDI, J.-M. BURKHARDT, G. CASIEZ, S. DONIKIAN, F. GOSSELIN, X. GRANIER, P. LE CALLET, V. LEPETIT, M. MARCHAL, G. MOREAU, J. PERRET, T. VIGIER.*Complexity and Scientific Challenges*, in "Virtual Reality and Augmented Reality - Myths and Realities", B. ARNALDI, P. GUITTON, G. MOREAU (editors), ISTE - Wiley, March 2018, vol. chapitre 3, p. 123-216 [*DOI :* 10.1002/9781119341031.CH3], https://hal.archives-ouvertes.fr/hal-01944915

[30] G. CASIEZ, X. GRANIER, M. HACHET, V. LEPETIT, G. MOREAU, O. NANNIPIERI.*Towards VE that are More Closely Related to the Real World*, in "Virtual Reality and Augmented Reality - Myths and Realities", B. ARNALDI, P. GUITTON, G. MOREAU (editors), ISTE - Wiley, March 2018, vol. chapitre 4 [*DOI :* 10.1002/9781119341031.CH4], https://hal.inria.fr/hal-01735022

Other Publications

[31] A. M. FEIT, M. NANCEL, D. WEIR, G. BAILLY, M. JOHN, A. KARRENBAUER, A. OULASVIRTA.*Élaboration de la disposition AZERTY modernisée*, June 2018, working paper or preprint, https://hal.inria.fr/hal-01826476

[32] M. NANCEL.*Historique et méthodologie de la nouvelle disposition de clavier AZERTY*, June 2018, working paper or preprint, https://hal.inria.fr/hal-01826384

References in notes

[33] 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

[34] 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

[35] D. C. ENGELBART.*Augmenting human intellect: a conceptual framework*, Stanford Research Institute, October 1962, n^{o} AFOSR-3233, http://www.dougengelbart.org/pubs/augment-3906.html

[36] N. FELLION, T. PIETRZAK, A. GIROUARD.*FlexStylus: leveraging flexion input for pen interaction*, in "Proceedings of UIST'17", ACM, October 2017, p. 375-385

[37] S. GREENBERG.*Toolkits and Interface Creativity*, in "Multimedia Tools Appl.", February 2007, vol. 32, n^{o} 2, p. 139–159, http://dx.doi.org/10.1007/s11042-006-0062-y

[38] J. GRUDIN.*A moving target: the evolution of Human-Computer Interaction*, in "The Human Computer Interaction handbook (3rd edition)", J. A. JACKO (editor), CRC Press, May 2012, p. xxvii-lxi, http://research.microsoft.com/en-us/um/people/jgrudin/publications/history/HCIhandbook3rd.pdf

[39] S. HUOT.*Designeering interaction: a missing link in the evolution of Human-Computer Interaction*, Université Paris-Sud, France, May 2013, 205 pages, Habilitation à diriger des recherches, https://hal.inria.fr/tel-00823763

[40] R. JOTA, A. NG, P. DIETZ, D. WIGDOR.*How fast is fast enough? A study of the effects of latency in direct-touch pointing tasks*, in "Proceedings of CHI'13", ACM, April 2013, p. 2291-2300, http://doi.acm.org/10.1145/2470654.2481317

[41] I. S. MACKENZIE, C. WARE.*Lag as a determinant of human performance in interactive systems*, in "Proceedings of CHI'93", ACM, April 1993, p. 488-493, http://doi.acm.org/10.1145/169059.169431

[42] W. E. MACKAY.*Responding to cognitive overload: coadaptation between users and technology*, in "Intellectica", 2000, vol. 30, n^{o} 1, p. 177-193, ARCo, http://intellectica.org/SiteArchives/archives/n30/30_06_Mackay.pdf

[43] N. MARQUARDT, S. GREENBERG.*Proxemic Interactions: From Theory to Practice*, Synthesis Lectures on Human-Centered Informatics, Morgan & Claypool, 2015, https://books.google.fr/books?id=2dPtBgAAQBAJ

[44] J. MATEJKA, W. LI, T. GROSSMAN, G. FITZMAURICE.*CommunityCommands: command recommendations for software applications*, in "Proceedings of UIST'09", ACM, October 2009, p. 193-202, http://dx.doi.org/10.1145/1622176.1622214

[45] M. NANCEL, E. PIETRIGA, O. CHAPUIS, M. BEAUDOUIN-LAFON.*Mid-air pointing on ultra-walls*, in "ACM ToCHI", October 2015, vol. 22, n^{o} 5, http://dx.doi.org/10.1145/2766448

[46] M. PANTIC, A. PENTLAND, A. NIJHOLT, T. S. HUANG.*Human computing and machine understanding of human behavior: a survey*, in "Artifical intelligence for human computing", T. S. HUANG, A. NIJHOLT, M.

PANTIC, A. PENTLAND (editors), LNCS, Springer, 2007, vol. 4451, p. 47-71, http://dx.doi.org/10.1007/978-3-540-72348-6_3

[47] V. RAMESH, C. HSU, M. AGRAWALA, B. HARTMANN.*ShowMeHow: Translating User Interface Instructions Between Applications*, in "Proceedings of UIST'11", ACM, October 2011, p. 127-134, http://doi.acm.org/10.1145/2047196.2047212

[48] J. SCARR, A. COCKBURN, C. GUTWIN, P. QUINN.*Dips and Ceilings: understanding and Supporting Transitions to Expertise in User Interfaces*, in "Proceedings of CHI'11", ACM, May 2011, p. 2741-2750, http://doi.acm.org/10.1145/1978942.1979348

[49] T. WINOGRAD, F. FLORES.*Understanding computers and cognition: a new foundation for design*, Addison-Wesley, May 1986, http://www.intellectbooks.co.uk/books/view-Book,id=4475/

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

1. **Team, Visitors, External Collaborators** **213**
2. **Overall Objectives** **214**
3. **Research Program** **215**
    3.1. Introduction 215
    3.2. Beyond Vectorial Models for NLP 215
    3.3. Adaptive Graph Construction 216
    3.4. Prediction on Graphs and Scalability 217
    3.5. Beyond Homophilic Relationships 218
4. **Application Domains** **219**
5. **Highlights of the Year** **219**
6. **New Software and Platforms** **220**
    6.1. CoRTeX 220
    6.2. Mangoes 220
    6.3. metric-learn 220
    6.4. MyLocalInfo 220
7. **New Results** **221**
    7.1. On the Bernstein-Hoeffding Method 221
    7.2. IncGraph: Incremental graphlet counting for topology optimisation 221
    7.3. Graph sampling with applications to estimating the number of pattern embeddings and the parameters of a statistical relational model 221
    7.4. A machine learning based framework to identify and classify long terminal repeat retrotransposons 221
    7.5. A Distributed Frank-Wolfe Framework for Learning Low-Rank Matrices with the Trace Norm 222
    7.6. Personalized and Private Peer-to-Peer Machine Learning 222
    7.7. Hiding in the Crowd: A Massively Distributed Algorithm for Private Averaging with Malicious Adversaries 222
    7.8. A Probabilistic Model for Joint Learning of Word Embeddings from Texts and Images 223
    7.9. A Framework for Understanding the Role of Morphology in Universal Dependency Parsing 223
    7.10. Online Reciprocal Recommendation with Theoretical Performance Guarantees 223
    7.11. On Similarity Prediction and Pairwise Clustering 223
    7.12. A Probabilistic Theory of Supervised Similarity Learning for Pointwise ROC Curve Optimization 224
    7.13. Escaping the Curse of Dimensionality in Similarity Learning: Efficient Frank-Wolfe Algorithm and Generalization Bounds 224
    7.14. Nonstochastic Bandits with Composite Anonymous Feedback 224
8. **Bilateral Contracts and Grants with Industry** **224**
    8.1. Coreference resolution 224
    8.2. Privacy preserving data mining for Mobility Data 225
    8.3. Predictive justice 225
9. **Partnerships and Cooperations** **225**
    9.1. Regional Initiatives 225
    9.2. National Initiatives 226
        9.2.1. ANR Pamela (2016-2020) 226
        9.2.2. ANR JCJC GRASP (2016-2020) 226
        9.2.3. ANR DEEP-Privacy (2019-2023) 226
        9.2.4. ANR-NFS REM (2016-2020) 226
        9.2.5. EFL (2010-2020) 227
    9.3. European Initiatives 227

9.3.1. FP7 & H2020 Projects 227
9.3.2. Collaborations in European Programs, Except FP7 & H2020 227
9.4. International Initiatives 227
9.4.1. Inria International Labs 227
9.4.2. Inria Associate Teams Not Involved in an Inria International Labs 228
9.5. International Research Visitors 228
9.5.1. Visits of International Scientists 228
9.5.2. Visits to International Teams 229
10. Dissemination **229**
10.1. Promoting Scientific Activities 229
10.1.1. Scientific Events Organisation 229
10.1.2. Scientific Events Selection 230
10.1.3. Journal 230
10.1.4. Invited Talks 230
10.1.5. Scientific Expertise 230
10.1.6. Research Administration 231
10.2. Teaching - Supervision - Juries 231
10.2.1. Teaching 231
10.2.2. Supervision 231
10.2.3. Juries 232
10.3. Popularization 232
10.3.1. Internal or external Inria responsibilities 232
10.3.2. Articles and contents 232
10.3.3. Interventions 232
11. Bibliography **233**

Project-Team MAGNET

Creation of the Team: 2013 January 01, updated into Project-Team: 2016 May 01

Keywords:

Computer Science and Digital Science:

A3.1. - Data
A3.1.3. - Distributed data
A3.1.4. - Uncertain data
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A3.5. - Social networks
A3.5.1. - Analysis of large graphs
A3.5.2. - Recommendation systems
A4.8. - Privacy-enhancing technologies
A9.4. - Natural language processing

Other Research Topics and Application Domains:

B1. - Life sciences
B1.1.10. - Systems and synthetic biology
B2. - Health
B2.2.4. - Infectious diseases, Virology
B2.3. - Epidemiology
B2.4.1. - Pharmaco kinetics and dynamics
B2.4.2. - Drug resistance
B5.10. - Biotechnology
B6.3. - Network functions
B7.1.2. - Road traffic
B8.3. - Urbanism and urban planning
B9.5.1. - Computer science
B9.5.4. - Chemistry
B9.5.6. - Data science
B9.6.8. - Linguistics
B9.6.10. - Digital humanities
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Aurelien Bellet [Inria, Researcher]
Pascal Denis [Inria, Researcher]
Claudio Gentile [Inria, Senior Researcher, from Oct 2018]

Jan Ramon [Inria, Senior Researcher]

Faculty Members

Bert Cappelle [Univ of Lille, Associate Professor, until Aug 2018]
Mikaela Keller [Univ of Lille, Associate Professor]
Marc Tommasi [Univ of Lille, Team leader, Professor, HDR]
Fabien Torre [Univ of Lille, Associate Professor, until Sep 2018]
Fabio Vitale [Univ of Lille, Associate Professor]
Joël Legrand [Univ of Lille, Associate Professor]

External Collaborator

Remi Gilleron [Univ of Lille, HDR]

Technical Staff

William de Vazelhes [Inria]
Arijus Pleska [Inria, until Sep 2018]
César Sabater [Inria]
Carlos Zubiaga Pena [Inria]

PhD Students

Mahsa Asadi [Inria, from Oct 2018]
Mathieu Dehouck [Univ des sciences et technologies de Lille]
Onkar Pandit [Inria]
Arijus Pleska [Inria, from Oct 2018]
Brij Mohan Lal Srivastava [Inria, from Oct 2018]
Mariana Vargas Vieyra [Inria]

Post-Doctoral Fellows

Melissa Ailem [Inria, until Sep 2018]
Thanh Le Van [Inria, until Jun 2018]
Bo Li [Univ of Lille]

Visiting Scientist

Tejas Kulkarni [Warwick University, from May 2018 until Aug 2018]

Administrative Assistant

Julie Jonas [Inria]

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 [37], [40].

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 [20], [42].

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 [41].

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 [20], [45]. 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 [42], [30]. 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 [32].

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 [31], [27], [29]). 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 [44]. We have already initiated some work on the coreference resolution with graph-based learning, by casting the problem as an instance of spectral clustering [29].

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 [39], face recognition [28], and text categorization [33].

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 ([46], [21], [22]). 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 ([23], [24]) 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 [35]. 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 [43].

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 [34], 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 [25]. 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 bandwidth 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 ([26], [36]) 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 [38]. 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. Domain 1

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.

We also target applications related to decentralized learning and privacy preserving systems when users or devices are interconnected in large networks. We develop solutions based on urban and mobility data where privacy is a specific requirement.

5. Highlights of the Year

5.1. Highlights of the Year

- Strengthening of the privacy aware machine learning activity with a new associate team with the Alan Turing Institute and the organization of a workshop at NeurIPS (formerly NIPS).
- New collaboration with Multispeech (Inria Nancy) on decentralized and private machine learning for speech processing leading to an ANR and an H2020 project.

5.1.1. Awards

AURÉLIEN BELLET received a best reviewer award (top 200 out of 3000) at the conference NeurIPS 2018. PASCAL DENIS received a Distinguished Senior Program Committee award at IJCAI-ECAI 2018.

6. New Software and Platforms

6.1. CoRTeX

Python library for noun phrase COreference Resolution in natural language TEXts
KEYWORD: Natural language processing
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 CONLL2012 and CONLLU annotation formats, and performing evaluation, notably 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. It currently supports use of the English or French language.

- Participant: Pascal Denis
- Partner: Orange Labs
- Contact: Pascal Denis
- URL: https://gitlab.inria.fr/magnet/CoRTeX

6.2. Mangoes

MAgnet liNGuistic wOrd vEctorS
KEYWORDS: Word embeddings - NLP
FUNCTIONAL DESCRIPTION: Process textual data and compute vocabularies and co-occurrence matrices. Input data should be raw text or annotated text. Compute word embeddings with different state-of-the art unsupervised methods. Propose statistical and intrinsic evaluation methods, as well as some visualization tools.

- Contact: Nathalie Vauquier
- URL: https://gitlab.inria.fr/magnet/mangoes

6.3. metric-learn

KEYWORDS: Machine learning - Python - Metric learning
FUNCTIONAL DESCRIPTION: Distance metrics are widely used in the machine learning literature. Traditionally, practicioners would choose a standard distance metric (Euclidean, City-Block, Cosine, etc.) using a priori knowledge of the domain. Distance metric learning (or simply, metric learning) is the sub-field of machine learning dedicated to automatically constructing optimal distance metrics.

This package contains efficient Python implementations of several popular metric learning algorithms.

- Partner: Parietal
- Contact: William De Vazelhes
- URL: https://github.com/metric-learn/metric-learn

6.4. MyLocalInfo

KEYWORDS: Privacy - Machine learning - Statistics
FUNCTIONAL DESCRIPTION: Decentralized algorithms for machine learning and inference tasks which (1) perform as much computation as possible locally and (2) ensure privacy and security by avoiding personal data leaves devices.

- Contact: Nathalie Vauquier
- URL: https://gitlab.inria.fr/magnet/mylocalinfo

7. New Results

7.1. On the Bernstein-Hoeffding Method

We consider extensions of Hoeffding's "exponential method" approach for obtaining upper estimates on the probability that a sum of independent and bounded random variables is significantly larger than its mean. We show that the exponential function in Hoeffding's approach can be replaced with any function which is non-negative, increasing and convex. As a result we generalize and improve upon Hoeffding's inequality. Our approach allows to obtain "missing factors" in Hoeffding's inequality. The later result is a rather weaker version of a theorem that is due to Michel Talagrand. Moreover, we characterize the class of functions with respect to which our method yields optimal concentration bounds. Finally, using ideas from the theory of Bernstein polynomials, we show that similar ideas apply under information on higher moments of the random variables ([4]).

7.2. IncGraph: Incremental graphlet counting for topology optimisation

Graphlets are small network patterns that can be counted in order to characterise the structure of a network (topology). As part of a topology optimisation process, one could use graphlet counts to iteratively modify a network and keep track of the graphlet counts, in order to achieve certain topological properties. Up until now, however, graphlets were not suited as a metric for performing topology optimisation; when millions of minor changes are made to the network structure it becomes computationally intractable to recalculate all the graphlet counts for each of the edge modifications. We propose IncGraph, a method for calculating the differences in graphlet counts with respect to the network in its previous state, which is much more efficient than calculating the graphlet occurrences from scratch at every edge modification made. In comparison to static counting approaches, our findings show IncGraph reduces the execution time by several orders of magnitude. The usefulness of this approach was demonstrated by developing a graphlet-based metric to optimise gene regulatory networks. IncGraph is able to quickly quantify the topological impact of small changes to a network, which opens novel research opportunities to study changes in topologies in evolving or online networks, or develop graphlet-based criteria for topology optimisation. IncGraph is freely available as an open-source R package on CRAN (incgraph). The development version is also available on GitHub (rcannood/incgraph) ([2]).

7.3. Graph sampling with applications to estimating the number of pattern embeddings and the parameters of a statistical relational model

Counting the number of times a pattern occurs in a database is a fundamental data mining problem. It is a subroutine in a diverse set of tasks ranging from pattern mining to supervised learning and probabilistic model learning. While a pattern and a database can take many forms, this paper focuses on the case where both the pattern and the database are graphs (networks). Unfortunately, in general, the problem of counting graph occurrences is #P-complete. In contrast to earlier work, which focused on exact counting for simple (i.e., very short) patterns, we present a sampling approach for estimating the statistics of larger graph pattern occurrences. We perform an empirical evaluation on synthetic and real-world data that validates the proposed algorithm, illustrates its practical behavior and provides insight into the trade-off between its accuracy of estimation and computational efficiency ([5]).

7.4. A machine learning based framework to identify and classify long terminal repeat retrotransposons

Transposable elements (TEs) are repetitive nucleotide sequences that make up a large portion of eukaryotic genomes. They can move and duplicate within a genome, increasing genome size and contributing to genetic diversity within and across species. Accurate identification and classification of TEs present in a genome is an important step towards understanding their effects on genes and their role in genome evolution. We introduce TE-LEARNER, a framework based on machine learning that automatically identifies TEs in a given genome and assigns a classification to them. We present an implementation of our framework towards LTR retrotransposons, a particular type of TEs characterized by having long terminal repeats (LTRs) at their boundaries. We evaluate the predictive performance of our framework on the well-annotated genomes of Drosophila melanogaster and Arabidopsis thaliana and we compare our results for three LTR retrotransposon superfamilies with the results of three widely used methods for TE identification or classification: REPEATMASKER, CENSOR and LTRDIGEST. In contrast to these methods, TE-LEARNER is the first to incorporate machine learning techniques, outperforming these methods in terms of predictive performance, while able to learn models and make predictions efficiently. Moreover, we show that our method was able to identify TEs that none of the above method could find, and we investigated TE-LEARNER's predictions which did not correspond to an official annotation. It turns out that many of these predictions are in fact strongly homologous to a known TE ([6]).

7.5. A Distributed Frank-Wolfe Framework for Learning Low-Rank Matrices with the Trace Norm

We consider the problem of learning a high-dimensional but low-rank matrix from a large-scale dataset distributed over several machines, where low-rankness is enforced by a convex trace norm constraint. We propose DFW-Trace, a distributed Frank-Wolfe algorithm which leverages the low-rank structure of its updates to achieve efficiency in time, memory and communication usage. The step at the heart of DFW-Trace is solved approximately using a distributed version of the power method. We provide a theoretical analysis of the convergence of DFW-Trace, showing that we can ensure sublinear convergence in expectation to an optimal solution with few power iterations per epoch. We implement DFW-Trace in the Apache Spark distributed programming framework and validate the usefulness of our approach on synthetic and real data, including the ImageNet dataset with high-dimensional features extracted from a deep neural network ([7]).

7.6. Personalized and Private Peer-to-Peer Machine Learning

The rise of connected personal devices together with privacy concerns call for machine learning algorithms capable of leveraging the data of a large number of agents to learn personalized models under strong privacy requirements. In this paper, we introduce an efficient algorithm to address the above problem in a fully decentralized (peer-to-peer) and asynchronous fashion, with provable convergence rate. We show how to make the algorithm differentially private to protect against the disclosure of information about the personal datasets, and formally analyze the trade-off between utility and privacy. Our experiments show that our approach dramatically outperforms previous work in the non-private case, and that under privacy constraints, we can significantly improve over models learned in isolation ([9]).

7.7. Hiding in the Crowd: A Massively Distributed Algorithm for Private Averaging with Malicious Adversaries

The amount of personal data collected in our everyday interactions with connected devices offers great opportunities for innovative services fueled by machine learning, as well as raises serious concerns for the privacy of individuals. In this paper, we propose a massively distributed protocol for a large set of users to privately compute averages over their joint data, which can then be used to learn predictive models. Our protocol can find a solution of arbitrary accuracy, does not rely on a third party and preserves the privacy of users throughout the execution in both the honest-but-curious and malicious adversary models. Specifically, we prove that the information observed by the adversary (the set of malicious users) does not significantly reduce the uncertainty in its prediction of private values compared to its prior belief. The level of privacy protection depends on a quantity related to the Laplacian matrix of the network graph and generally improves with the size of the graph. 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 ([15]).

7.8. A Probabilistic Model for Joint Learning of Word Embeddings from Texts and Images

Several recent studies have shown the benefits of combining language and perception to infer word embeddings. These multimodal approaches either simply combine pre-trained textual and visual representations (e.g. features extracted from convolutional neural networks), or use the latter to bias the learning of textual word embeddings. In this work, we propose a novel probabilistic model to formalize how linguistic and perceptual inputs can work in concert to explain the observed word-context pairs in a text corpus. Our approach learns textual and visual representations jointly: latent visual factors couple together a skip-gram model for co-occurrence in linguistic data and a generative latent variable model for visual data. Extensive experimental studies validate the proposed model. Concretely, on the tasks of assessing pairwise word similarity and image/caption retrieval, our approach attains equally competitive or stronger results when compared to other state-of-the-art multimodal models ([8]).

7.9. A Framework for Understanding the Role of Morphology in Universal Dependency Parsing

We present a simple framework for characterizing morphological complexity and how it encodes syntactic information. In particular, we propose a new measure of morpho-syntactic complexity in terms of governor-dependent preferential attachment that explains parsing performance. Through experiments on dependency parsing with data from Universal Dependencies (UD), we show that representations derived from morphological attributes deliver important parsing performance improvements over standard word form embeddings when trained on the same datasets. We also show that the new morpho-syntactic complexity measure is predictive of the gains provided by using morphological attributes over plain forms on parsing scores, making it a tool to distinguish languages using morphology as a syntactic marker from others ([11]).

7.10. Online Reciprocal Recommendation with Theoretical Performance Guarantees

A reciprocal recommendation problem is one where the goal of learning is not just to predict a user's preference towards a passive item (e.g., a book), but to recommend the targeted user on one side another user from the other side such that a mutual interest between the two exists. The problem thus is sharply different from the more traditional items-to-users recommendation, since a good match requires meeting the preferences at both sides. We initiate a rigorous theoretical investigation of the reciprocal recommendation task in a specific framework of sequential learning. We point out general limitations, formulate reasonable assumptions enabling effective learning and, under these assumptions, we design and analyze a computationally efficient algorithm that uncovers mutual likes at a pace comparable to that achieved by a clairvoyant algorithm knowing all user preferences in advance. Finally, we validate our algorithm against synthetic and real-world datasets, showing improved empirical performance over simple baselines ([13]).

7.11. On Similarity Prediction and Pairwise Clustering

We consider the problem of clustering a finite set of items from pairwise similarity information. Unlike what is done in the literature on this subject, we do so in a passive learning setting, and with no specific constraints on the cluster shapes other than their size. We investigate the problem in different settings: i. an online setting, where we provide a tight characterization of the prediction complexity in the mistake bound model, and ii. a standard stochastic batch setting, where we give tight upper and lower bounds on the achievable generalization error. Prediction performance is measured both in terms of the ability to recover the similarity function encoding the hidden clustering and in terms of how well we classify each item within the set. The proposed algorithms are time efficient ([12]).

7.12. A Probabilistic Theory of Supervised Similarity Learning for Pointwise ROC Curve Optimization

The performance of many machine learning techniques depends on the choice of an appropriate similarity or distance measure on the input space. Similarity learning (or metric learning) aims at building such a measure from training data so that observations with the same (resp. different) label are as close (resp. far) as possible. In this paper, similarity learning is investigated from the perspective of pairwise bipartite ranking, where the goal is to rank the elements of a database by decreasing order of the probability that they share the same label with some query data point, based on the similarity scores. A natural performance criterion in this setting is pointwise ROC optimization: maximize the true positive rate under a fixed false positive rate. We study this novel perspective on similarity learning through a rigorous probabilistic framework. The empirical version of the problem gives rise to a constrained optimization formulation involving U-statistics, for which we derive universal learning rates as well as faster rates under a noise assumption on the data distribution. We also address the large-scale setting by analyzing the effect of sampling-based approximations. Our theoretical results are supported by illustrative numerical experiments ([14]).

7.13. Escaping the Curse of Dimensionality in Similarity Learning: Efficient Frank-Wolfe Algorithm and Generalization Bounds

Similarity and metric learning provides a principled approach to construct a task-specific similarity from weakly supervised data. However, these methods are subject to the curse of dimensionality: as the number of features grows large, poor generalization is to be expected and training becomes intractable due to high computational and memory costs. In this paper, we propose a similarity learning method that can efficiently deal with high-dimensional sparse data. This is achieved through a parameterization of similarity functions by convex combinations of sparse rank-one matrices, together with the use of a greedy approximate Frank-Wolfe algorithm which provides an efficient way to control the number of active features. We show that the convergence rate of the algorithm, as well as its time and memory complexity, are independent of the data dimension. We further provide a theoretical justification of our modeling choices through an analysis of the generalization error, which depends logarithmically on the sparsity of the solution rather than on the number of features. Our experiments on datasets with up to one million features demonstrate the ability of our approach to generalize well despite the high dimensionality as well as its superiority compared to several competing methods ([16]).

7.14. Nonstochastic Bandits with Composite Anonymous Feedback

We investigate a nonstochastic bandit setting in which the loss of an action is not immediately charged to the player, but rather spread over at most d consecutive steps in an adversarial way. This implies that the instantaneous loss observed by the player at the end of each round is a sum of as many as d loss components of previously played actions. Hence, unlike the standard bandit setting with delayed feedback, here the player cannot observe the individual delayed losses, but only their sum. Our main contribution is a general reduction transforming a standard bandit algorithm into one that can operate in this harder setting. We also show how the regret of the transformed algorithm can be bounded in terms of the regret of the original algorithm. Our reduction cannot be improved in general: we prove a lower bound on the regret of any bandit algorithm in this setting that matches (up to log factors) the upper bound obtained via our reduction. Finally, we show how our reduction can be extended to more complex bandit settings, such as combinatorial linear bandits and online bandit convex optimization ([10]).

8. Bilateral Contracts and Grants with Industry

8.1. Coreference resolution

Along a collaboration with Orange, we developed a Natural Language Processing library for co-reference resolution. The library is based on a previous work (CorTeX) and was extended in several ways. It handles the French language, it includes new features based on vectorial representations of words (word embeddings) and it is more scalable. PASCAL DENIS is the local PI at Inria of this project.

8.2. Privacy preserving data mining for Mobility Data

JAN RAMON is the local PI at Inria for the ADEME-MUST project (Méthodologie d'exploitation des données d'usage des véhicules et d'identification de nouveaux services pour les usagers et les territoires). We study machine learning and data mining methods for knowledge discovery from mobility data, which are time-stamped signals collected from cars, for example, GPS locations, accelerations and fuel consumption. We aim to discover knowledge that helps us to address important questions in the transportation system such as road safety, traffic congestion, parking, ride-sharing, pollution and energy consumption. As the mobility data contains a lot of personal information, for instance, driving styles and locations of the users, we hence also study methods that allow the users to keep their personal data and only exchange part of them to collaboratively derive the knowledge.

The project has four partners, including, Xee company, CEREMA, i-Trans and Inria. The Xee company is responsible for recruiting drivers and collecting the data. CEREMA and i-Trans function as domain experts who help us to form the questions and verify the analytical results. MAGNET is responsible for developing and applying data mining methods for analyzing the data. The developed methods and the discovered knowledge from the project will be transferred to Metropole Lille and ADEME.

8.3. Predictive justice

Claim assistance is a French company that develops assistance for conflict resolution. The main service is RefundMyTicket [0]. In the general project of partial automation of analysis of complains, we have provided consulting and supervision. The general approach was to be able to analyze, parse and reason on legal texts. We have developed strategies based on natural language processing in the specific domain of legal texts. Techniques include learning representation and structured prediction among others.

9. Partnerships and Cooperations

9.1. Regional Initiatives

We conducted research in collaboration with J. Senechal from the department of law in Lille University. We are interested in studying the impact of technological choices regarding computation models in the perspective of the GDPR.

We strengthened our partnership with the linguistic laboratory STL in Lille university. We have welcomed Bert Cappelle for a stay (delegation) in the group. The topic of this collaboration was to study modal verbs and the translation of the notion of compositionality when applied to vectorial representation of words.

We initiated a collaboration with cognitive scientists (Angèle Brunellière and Jérémie Jozefowiez) from the psychology department, which resulted in a submission to a multidisciplinary Huma-Num project, to be funded by the Réseau National des Maisons des Sciences de l'Homme (RNMSH).

We started working with Christopher Fletcher (CNRS) from the History department.

These collaborations heavily rely on our work on distributional semantics and word embeddings to provide new insights into these different fields, hence also on the Mangoes toolkit developed in the team.

[0] https://www.refundmyticket.net

We participate to the *Data Advanced data science and technologies* project (CPER Data). This project 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. The project MyLocalInfo is an open API for privacy-friendly collaborative computing in the internet of things.

9.2. National Initiatives

9.2.1. ANR Pamela (2016-2020)

Participants: MARC TOMMASI [correspondent], AURÉLIEN BELLET, RÉMI GILLERON, JAN RAMON, MAHSA ASADI

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 DEEP-Privacy (2019-2023)

Participants: MARC TOMMASI [correspondent], AURÉLIEN BELLET, PASCAL DENIS, JAN RAMON, BRIJ SRIVASTAVA

DEEP-PRIVACY proposes a new paradigm based on a distributed, personalized, and privacy-preserving approach for speech processing, with a focus on machine learning algorithms for speech recognition. To this end, we propose to rely on a hybrid approach: the device of each user does not share its raw speech data and runs some private computations locally, while some cross-user computations are done by communicating through a server (or a peer-to-peer network). To satisfy privacy requirements at the acoustic level, the information communicated to the server should not expose sensitive speaker information.

9.2.4. ANR-NFS REM (2016-2020)

Participants: PASCAL DENIS [correspondent], BO LI

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.

9.2.5. 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

Program: H2020 ICT-29-2018 (RIA)

Project acronym: COMPRISE

Project title: Cost-effective, Multilingual, Privacy-driven voice-enabled Services

Duration: Dec 2018- Nov 2021

Coordinator: Emmanuel Vincent

Other partners: Inria Multispeech, Ascora GmbH, Netfective Technology SA, Rooter Analysis SL, Tilde SIA, University of Saarland

Participants: AURÉLIEN BELLET, MARC TOMMASI, BRIJ SRIVASTAVA

Abstract: COMPRISE will define a fully private-by-design methodology and tools that will reduce the cost and increase the inclusiveness of voice interaction technologies.

9.3.2. Collaborations in European Programs, Except FP7 & H2020

9.3.2.1. 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)

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. TextLink will enhance the experience and performance of human translators, lexicographers, language technology and language learners alike.

9.4. International Initiatives

9.4.1. Inria International Labs

Inria@SiliconValley
Associate Team involved in the International Lab:

9.4.1.1. LEGO

Title: LEarning GOod representations for natural language processing
International Partner (Institution - Laboratory - Researcher):
USC (United States), Prof. Fei Sha.
Start year: 2016
See also: https://team.inria.fr/lego/
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.

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

North-European Associate Team PAD-ML: Privacy-Aware Distributed Machine Learning.
International Partner: the PPDA team at the Alan Turing Institute.
Start year: 2018
In the context of increasing legislation on data protection (e.g., the recent GDPR), an important challenge is to develop privacy-preserving algorithms to learn from datasets distributed across multiple data owners who do not want to share their data. The goal of this joint team is to devise novel privacy-preserving, distributed machine learning algorithms and to assess their performance and guarantees in both theoretical and practical terms.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Tejas Kulkarni (University of Warwick) visited the team from May to August 2018 to work with Aurélien Bellet, Marc Tommasi and Jan Ramon on privacy-preserving computation of U-statistics.
- Larisa Soldatova (Brunel University) visited the team in June 2018 to work with Jan Ramon on probabilistic reasoning for biomedical applications.
- Raouf Kerkouche (Inria Privatics) visited the team for 2 weeks in July 2018 to work with Aurélien Bellet and Marc Tommasi on federated and decentralized learning from medical data.
- Guillaume Rabusseau (Université de Montréal) visited the team for 1 week in July 2018 to work with Aurélien Bellet and Marc Tommasi on multi-task distributed spectral learning.
- Daphner Ezer, Adrià Gascón, Matt Kusner, Brooks Paige (all from Alan Turing Institute) and Hamed Haddadi (Imperial College London) visited the team for 2 days in October 2018 for the kick-off of the PAD-ML associate team.

Several international researchers have also been invited to give a talk at the MAGNET seminar:

- D. Hovy (Bocconi Univ.): Retrofit Everything: Injecting External Knowledge into Neural Networks to Gain Insights from Big Data.
- A. Trask (OpenMined): OpenMined - Building Tools for Safe AI.
- C. Biemann (Univ. Hamburg): Adaptive Interpretable Language Technology.
- W. Daelemans (Univ. Antwerp): Profiling authors from social media texts.

9.5.1.1. Internships

- Igor Axinti explored several ways to compare word embeddings and studied the minimal corpus size for the comparison to be meaningful. He applied some of his findings to comparing two corpus in middle french from the 15th century, one originating from London and the other from Flanders. He produced a querying interface to allow Christopher Fletcher (IRHiS), who provided the data, explore and compare the embeddings spaces.
- Nicolas Crosetti (joint internship with Joachim Niehren and Florent Cappelli, Links) worked on dependency-weighted aggregation, i.e., aggregation where the elements to aggregate are weighted according to the extent where they correspond to independent observations.
- Arthur d'Azemar worked on decentralized recommender systems in collaboration with the WIDE team in Inria Rennes (François Taïani). Arthur has applied metric learning techniques in order to learn a K-nn graph for personalized and adaptive user-based recommendations.
- Antoine Capriski worked on the analysis of word semantic change in political texts in collaboration with Caroline Le Pennec (UC Berkeley). He used the techniques of word embeddings to analyze of corpus of political manifestos from the French general elections for the period 1958-1993.
- Most of the works on machine learning and privacy make the assumption that learners are honest but curious. Alexandre Huat worked on making protocols for private machine learning more robust again malicious attacks.

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

- FABIO VITALE is on leave at Department of Computer Science of Sapienza University (Rome, Italy) in the Algorithms Randomization Computation group with Prof. Alessandro Panconesi and Prof. Flavio Chierichetti. His current work on machine learning in graphs follows three directions:
  - designing new online reciprocal recommenders analyzing their performance both in theory and in practice,
  - clustering a finite set of items from pairwise similarity information in different learning settings,
  - introducing a new online learning framework encompassing several problems where the environment changes over time, and an efficient and very scalable unifying approach to solve the related general learning problem.

  Current (and unfinished) ongoing research also includes the following topics: low-stretch spanning trees, active learning in correlation clustering problems, hierarchical clustering.
- AURÉLIEN BELLET visited the Alan Turing Institute (London) and Amazon Research Cambridge for 1 week in February 2018. He worked with Adrià Gascón and Borja Balle on privacy-preserving machine learning.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- AURÉLIEN BELLET was a member of the organization committee of the PPML workshop at NeurIPS'18. [0] The workshop was on Privacy Preserving Machine Learning and had among its invited speakers Shafi Goldwasser (Gödel and Turing Prize), Adam Smith (Gödel Prize).
- AURÉLIEN BELLET co-organized the kick-off workshop of the associated team PAD-ML with the Alan Turing Institute. [0] The workshop was held at Inria Lille and featured speakers from MAGNET and the Alan Turing Institute.

[0] https://neurips.cc/Conferences/2018/Schedule?showEvent=10934
[0] https://team.inria.fr/magnet/workshop-on-privacy-aware-distributed-machine-learning/

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- AURÉLIEN BELLET served as PC member for AISTATS'19, ICML'18, NIPS'18, IJCAI'18 Sister Conference, PiMLAI workshop at ICML'18, and CAP'18.
- PASCAL DENIS served as PC member for ACL'18, CONLL'18, EMNLP'18, NAACL'18, NIPS'18, IJCAI-ECAI'18 (Senior PC), CRAC Workshop at NAACL'18.
- MARC TOMMASI served as PC member for AAAI'18, ICML'18, CAP'18, IJCAI'18 (Senior PC chair), AISTATS'18, NIPS'18.
- JAN RAMON served as PC member for AAAI'19, AISTATS'19, IEEE-BigData'18, CIKM'18, DS'18, ECML/PKDD'18, EKAW'18, IEEE-ICDM'18, ICML'18, ILP'18, LOD'18, MLG'18, NIPS'18, SDM'18, TDLSG'18.
- MIKAELA KELLER served as PC member for ICML'18, CAP'18.
- RÉMI GILLERON served as PC member for NIPS'18, CAP'18, AISTATS'19 and ICLR'19.

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- AURÉLIEN BELLET was reviewer for Machine Learning Journal and IEEE/ACM Transactions on Networking.
- PASCAL DENIS was reviewer for Computational Linguistics, IJCAI-ECAI Surveys, and Language Resources and Evaluation.
- JAN RAMON was member of the editorial boards of Machine Learning Journal (MLJ) and Data Mining and Knowledge Discovery (DMKD). JAN RAMON was reviewer for among others JMLR, TPAMI, JIIS.

10.1.4. Invited Talks

- AURÉLIEN BELLET gave invited talks at the EPFL-Inria 2018 workshop [0] and the Journées de Statistique 2018 (session SSFAM). [0]
- AURÉLIEN BELLET was invited to talk at the seminars of Inria WIDE, Télécom ParisTech, Statistics Seminar of Paris 6/7, CMLA (ENS Paris Saclay) and Naver Labs Europe.
- PASCAL DENIS gave an invited talk at the Séminaire Langage, SCALab, Université de Lille, 26/01/18.

10.1.5. Scientific Expertise

- AURÉLIEN BELLET was a member of the jury for the Gilles-Kahn PhD award of the French Society of Computer Science (SIF), sponsored by the French Academy of Sciences. [0]
- AURÉLIEN BELLET acted as external reviewer for the French National Research Agency (ANR), track "Projets de Recherche Collaborative – International".
- JAN RAMON was an external reviewer for the Swiss National Science Foundation (SNF).
- JAN RAMON was an external reviewer for the Vienna Science and Technology Fund (WWTF).
- JAN RAMON acted as an expert for the H2020 CoE and IMI programs.

[0] https://project.inria.fr/epfl-Inria/workshops/workshop-2018/
[0] http://jds2018.sfds.asso.fr/
[0] https://www.societe-informatique-de-france.fr/recherche/prix-de-these-gilles-kahn/

10.1.6. Research Administration

- MIKAELA KELLER is member of the Conseil du laboratoire CRIStAL.
- FABIEN TORRE is member of the bureau du Conseil National des Universités (section 27).
- PASCAL DENIS served as a member of the CNRS Pre-GDR NLP Group.
- PASCAL DENIS was elected to Comité National du CNRS, section 34 (Sciences du Langage).

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Licence SHS: JOËL LEGRAND, Traitement de textes et tableur, 10h, L1, Université Lille.

Licence SHS: MARC TOMMASI, Langages du Web, 24h, L2, Université Lille.

Licence MIASHS: MIKAELA KELLER, Python 1, 40h, L1, Université Lille.

Licence MIASHS: MARC TOMMASI, Codage et représentation de l'information, 48h, L1, Université Lille.

Licence MIASHS: MIKAELA KELLER, Codage et représentation de l'information, 42h, L1, Université Lille.

Licence SoQ (SHS): MIKAELA KELLER, Algorithmique de graphes, 24h, L3, Université Lille.

Licence MARC TOMMASI C2i 12h, Université Lille.

Licence MARC TOMMASI Humanités numériques - Découvrir et faire découvrir la programmation, 20h, Université Lille/

Master MIASHS: MIKAELA KELLER, Algorithmes fondamentaux de la fouille de données, 60h, M1, Université Lille.

Master MIASHS: JOËL LEGRAND, Apprentissage et émergence de comportements, 30h, M2, Université Lille.

Master Data Analysis & Decision Making: AURÉLIEN BELLET, Machine Learning, 12h, Ecole Centrale de Lille.

Master / Master Spécialisé Big Data: AURÉLIEN BELLET, Advanced Machine Learning, 15h, Télécom ParisTech.

Formation continue (Certificat d'Études Spécialisées Data Scientist): AURÉLIEN BELLET, Supervised Learning and Support Vector Machines, 17.5h, Télécom ParisTech.

Master Informatique: PASCAL DENIS, Fondements de l'Apprentissage Automatique, 46h, M1, Université de Lille.

10.2.2. Supervision

Postdoc: MELISSA AILEM, InriaSiliconValley postdoctoral grant, supervised by AURÉLIEN BELLET, MARC TOMMASI, PASCAL DENIS and FEI SHA (University of Southern California).

Postdoc: BO LI, supervised by PASCAL DENIS on ANR REM, Model Sense Disambiguation, since December 2017.

PhD: GÉRAUD LE FALHER, Characterizing edges in signed and vector-valued graphs. April 16th 2018, MARC TOMMASI and FABIO VITALE and CLAUDIO GENTILE.

Phd: ASHRAF M. KIBRIYA, Mining Frequent Patterns in Large Networks, June 2018, JAN RAMON.

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: ONKAR PANDIT, Graph-based Semi-supervised Linguistic Structure Prediction, since Dec. 2017, PASCAL DENIS, MARC TOMMASI and LIVA RALAIVOLA (University of Marseille).

PhD in progress: Mariana Vargas Vieyra, Adaptive Graph Learning with Applications to Natural Language Processing, since Jan. 2018. Pascal Denis and Aurélien Bellet and Marc Tommasi.

PhD in progress: Brij Srivastava, Representation Learning for Privacy-Preserving Speech Recognition, since Oct 2018 Aurélien Bellet and Marc Tommasi and Emmanuel Vincent.

PhD in progress: Mahsa Asadi, On Decentralized Machine Learning, since Oct 2018. Aurélien Bellet and Marc Tommasi.

PhD in progress: Nicolas Crosetti, Privacy Risks of Aggregates in Data Centric-Workflows, since Oct 2018. Florent Capelli and Sophie Tison and Joachim Niehren and Jan Ramon.

PhD in progress: Robin Vogel, Learning to rank by similarity and performance optimization in biometric identification, since 2017 (CIFRE thesis with IDEMIA and Télécom ParisTech). Aurélien Bellet, Stéphan Clémençon and Anne Sabourin.

10.2.3. Juries

- Aurélien Bellet was member of the PhD jury of Guillaume Papa (Télécom ParisTech), Wenjie Zheng (Sorbonne Université), Michael Blot (Sorbonne Université).
- Marc Tommasi was member of the Phd jury of Gaëtan Hadjeres (*Rapporteur*), Alexandre Bérard (*Head*), Olivier Ruas (*Rapporteur*), Valentina Zantedeschi.
- Pascal Denis was *rapporteur* on the Phd jury of Elena Knyazeva, Université Paris-Saclay.
- Mikaela Keller was member of the recruitment committee for Assistant Professors in Computer Science at Université of Lille and at Université de St-Étienne.
- Mikaela Keller was member of the Phd jury of Damien Fourure (Université de St-Étienne) and of the HDR jury of Renaud Lopes (CHRU Lille).
- Rémi Gilleron was head of the PhD jury of Romain Warlop (Université de Lille).
- Pascal Denis was a member of hiring committee for Junior Research Scientist at Inria Lille.
- Marc Tommasi was member of the recruitment committee Assistant Professors in Computer Science at Université of Lille and for professor position at INSA de Lyon.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

- Aurélien Bellet is the scientific mediation contact for Inria Lille center.
- Pascal Denis served as committee member on the Inria Lille Commission Emploi Recherche (CER).
- Pascal Denis also served as committee member on Commission de Développement Technologique (CDT).
- Pascal Denis is administrator of Inria membership to Linguistic Data Consortium (LDC).

10.3.2. Articles and contents

- Aurélien Bellet and Marc Tommasi provided expertise for an upcoming TV program on Arte about new technologies.

10.3.3. Interventions

- National events: Jan Ramon and Marc Tommasi participate to a round-table meeting at the *Fête des libertés numériques* for the RGPD day [0].
- In educational institutions: Marc Tommasi gave a talk on privacy and machine learning in Journées polytech [0].

[0] https://www.meshs.fr/page/donnees_personnelles_et_droits_et_libertes_numeriques
[0] http://www.polytech-lille.fr/big-data-machine-mearning-p11419.html#.WqlHjExFxPb

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] G. LE FALHER.*Characterizing Edges in Signed and Vector-Valued Graphs*, Université de Lille, April 2018, https://hal.inria.fr/tel-01824215

Articles in International Peer-Reviewed Journal

[2] R. CANNOODT, J. RUYSSINCK, J. RAMON, K. DE PRETER, Y. SAEYS.*IncGraph: Incremental graphlet counting for topology optimisation*, in "PLoS ONE", April 2018, vol. 13, n^{o} 4 [*DOI :* 10.1371/ JOURNAL.PONE.0195997], https://hal.inria.fr/hal-01814675

[3] B. CAPPELLE, P. DENIS, M. KELLER.*Facing the facts of fake: a distributional semantics and corpus annotation approach*, in "Yearbook of the German Cognitive Linguistics Association", November 2018, https://hal.archives-ouvertes.fr/hal-01959609

[4] C. PELEKIS, J. RAMON, Y. WANG.*On the Bernstein-Hoeffding method*, in "Bulletin of the Hellenic Mathematical Society", June 2018, vol. 62, p. 31-43, https://hal.inria.fr/hal-01814651

[5] I. RAVKIC, M. ZNIDARŠIČ, J. RAMON, J. DAVIS.*Graph sampling with applications to estimating the number of pattern embeddings and the parameters of a statistical relational model*, in "Data Mining and Knowledge Discovery", 2018, 36, https://hal.inria.fr/hal-01725971

[6] L. SCHIETGAT, C. VENS, J. RAMON, R. CERRI, C. N. FISCHER, E. D. COSTA, C. M. A. CARARETO, H. BLOCKEEL.*A machine learning based framework to identify and classify long terminal repeat retrotransposons*, in "PLoS Computational Biology", April 2018, vol. 14, n^{o} 4, p. 1-21 [*DOI :* 10.1371/JOURNAL.PCBI.1006097], https://hal.inria.fr/hal-01814669

[7] W. ZHENG, A. BELLET, P. GALLINARI.*A Distributed Frank-Wolfe Framework for Learning Low-Rank Matrices with the Trace Norm*, in "Machine Learning", 2018, https://hal.inria.fr/hal-01922994

International Conferences with Proceedings

[8] M. AILEM, B. ZHANG, A. BELLET, P. DENIS, F. SHA.*A Probabilistic Model for Joint Learning of Word Embeddings from Texts and Images*, in "Conference on Empirical Methods in Natural Language Processing (EMNLP 2018)", Brussels, Belgium, 2018, https://hal.inria.fr/hal-01922985

[9] A. BELLET, R. GUERRAOUI, M. TAZIKI, M. TOMMASI.*Personalized and Private Peer-to-Peer Machine Learning*, in "AISTATS 2018 - 21st International Conference on Artificial Intelligence and Statistics", Lanzarote, Spain, April 2018, p. 1-20, https://arxiv.org/abs/1705.08435 , https://hal.inria.fr/hal-01745796

[10] N. CESA-BIANCHI, C. GENTILE, Y. MANSOUR.*Nonstochastic Bandits with Composite Anonymous Feedback*, in "COLT 2018 - 31st Annual Conference on Learning Theory", Stockholm, Sweden, July 2018, vol. 75, p. 1 - 23, https://hal.inria.fr/hal-01916981

[11] M. DEHOUCK, P. DENIS.*A Framework for Understanding the Role of Morphology in Universal Dependency Parsing*, in "EMNLP 2018 - Conference on Empirical Methods in Natural Language Processing", Brussels, Belgium, Proceedings of EMNLP 2018, October 2018, https://hal.archives-ouvertes.fr/hal-01943934

[12] S. PASTERIS, F. VITALE, C. GENTILE, M. HERBSTER.*On Similarity Prediction and Pairwise Clustering*, in "ALT 2018 - 29th International Conference on Algorithmic Learning Theory", Lanzarote, Spain, April 2018, vol. 83, p. 1 - 28, https://hal.inria.fr/hal-01916976

[13] F. VITALE, N. PAROTSIDIS, C. GENTILE.*Online Reciprocal Recommendation with Theoretical Performance Guarantees*, in "NIPS 2018 - 32nd Conference on Neural Information Processing Systems", Montreal, Canada, December 2018, https://hal.inria.fr/hal-01916979

[14] R. VOGEL, A. BELLET, S. CLÉMENÇON.*A Probabilistic Theory of Supervised Similarity Learning for Pointwise ROC Curve Optimization*, in "International Conference on Machine Learning (ICML 2018)", Stockholm, Sweden, 2018, https://hal.inria.fr/hal-01922988

Research Reports

[15] P. DELLENBACH, A. BELLET, J. RAMON.*Hiding in the Crowd: A Massively Distributed Algorithm for Private Averaging with Malicious Adversaries*, Inria, 2018, https://hal.inria.fr/hal-01923000

[16] K. LIU, A. BELLET.*Escaping the Curse of Dimensionality in Similarity Learning: Efficient Frank-Wolfe Algorithm and Generalization Bounds*, Inria, 2018, https://hal.inria.fr/hal-01923006

Other Publications

[17] F. CAPELLI, N. CROSETTI, J. NIEHREN, J. RAMON.*Dependency Weighted Aggregation on Factorized Databases*, January 2019, https://arxiv.org/abs/1901.03633 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01981553

[18] O.-A. MAILLARD, M. ASADI.*Upper Confidence Reinforcement Learning exploiting state-action equivalence*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01945034

[19] J. RAMON.*Exploiting traffic data*, December 2018, Presentation given at Inria Meetup (Lille), https://hal.inria.fr/hal-01879941

References in notes

[20] 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

[21] 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

[22] M. BELKIN, P. NIYOGI.*Towards a Theoretical Foundation for Laplacian-Based Manifold Methods*, in "Journal of Computer and System Sciences", 2008, vol. 74, n° 8, p. 1289-1308

[23] A. Bellet, A. Habrard, M. Sebban.*A Survey on Metric Learning for Feature Vectors and Structured Data*, in "CoRR", 2013, vol. abs/1306.6709

[24] A. Bellet, A. Habrard, M. Sebban.*Metric Learning*, Morgan & Claypool Publishers, 2015

[25] 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

[26] P. Blau.*Inequality and Heterogeneity: A Primitive Theory of Social Structure*, MACMILLAN Company, 1977, http://books.google.fr/books?id=jvq2AAAAIAAJ

[27] 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

[28] 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

[29] 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

[30] D. Das, S. Petrov.*Unsupervised Part-of-Speech Tagging with Bilingual Graph-Based Projections*, in "ACL", 2011, p. 600-609

[31] 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

[32] E. R. Fernandes, U. Brefeld.*Learning from Partially Annotated Sequences*, in "ECML/PKDD", 2011, p. 407-422

[33] 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

[34] 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

[35] M. Gomez-Rodriguez, J. Leskovec, A. Krause.*Inferring networks of diffusion and influence*, in "Proc. of KDD", 2010, p. 1019-1028

[36] 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

[37] A. NENKOVA, K. MCKEOWN.*A Survey of Text Summarization Techniques*, in "Mining Text Data", Springer, 2012, p. 43-76

[38] 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

[39] 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

[40] 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

[41] 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

[42] A. SUBRAMANYA, S. PETROV, F. C. N. PEREIRA.*Efficient Graph-Based Semi-Supervised Learning of Structured Tagging Models*, in "EMNLP", 2010, p. 167-176

[43] 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

[44] 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

[45] K. K. YUZONG LIU.*Graph-Based Semi-Supervised Learning for Phone and Segment Classification*, in "Proceedings of Interspeech", Lyon, France, 2013

[46] X. ZHU, Z. GHAHRAMANI, J. LAFFERTY.*Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions*, in "Proc. of ICML", 2003, p. 912-919

Team MEPHYSTO-POST

Quantitative methods for stochastic models in physics

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

1. Team, Visitors, External Collaborators **241**
2. Overall Objectives **241**
3. Research Program **242**
3.1. Time asymptotics: Stationary states, solitons, and stability issues 242
3.2. Derivation of macroscopic laws from microscopic dynamics 242
3.3. Numerical methods: analysis and simulations 243
4. New Results **243**
4.1. Exponential time-decay for discrete Fokker–Planck equations 243
4.2. Energy preserving methods for nonlinear Schrödinger equations 243
4.3. Diffusive and superdiffusive behavior in one-dimensional chains of oscillators 243
4.4. Microscopic description of moving interfaces 244
4.5. Stability analysis of a Vlasov-Wave system 244
4.6. Orbital stability in the presence of symmetries 245
4.7. Measuring nonclassicality of bosonic field quantum state 245
4.8. The Cauchy problem for the Landau–Lifshitz–Gilbert equation in BMO and self-similar solutions 245
4.9. The Sine–Gordon regime of the Landau–Lifshitz equation with a strong easy-plane anisotropy 245
4.10. Mutual information of wireless channels and block-Jacobi ergodic operators 245
4.11. DLR equations and rigidity for the Sine-beta process 246
4.12. Time-frequency transforms of white noises and Gaussian analytic functions 246
4.13. Energy of the Coulomb gas on the sphere at low temperature 246
4.14. Polynomial ensembles and recurrence coefficients 246
4.15. Concentration for Coulomb gases and Coulomb transport inequalities 246
5. Partnerships and Cooperations **247**
5.1. National Initiatives 247
5.2. European Initiatives 247
5.3. International Research Visitors 247
6. Dissemination **247**
6.1. Promoting Scientific Activities 247
6.1.1. Scientific Events Organisation 247
6.1.2. Journal 248
6.1.3. Invited Talks 248
6.1.4. Research Administration 248
6.2. Teaching - Supervision - Juries 248
6.2.1. Teaching 248
6.2.2. Supervision 248
6.2.3. Juries 248
6.3. Popularization 249
7. Bibliography **249**

Team MEPHYSTO-POST

Creation of the Team: 2017 October 01

Keywords:

Computer Science and Digital Science:

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.4. - Multiscale modeling
A6.2.1. - Numerical analysis of PDE and ODE

Other Research Topics and Application Domains:

B9.5.2. - Mathematics

1. Team, Visitors, External Collaborators

Research Scientists

Guillaume Dujardin [Team leader, Inria Researcher, HDR]
Marielle Simon [Inria Researcher]

Faculty Members

Denis Bonheure [Université Libre de Bruxelles, Professor, until Mar 2018]
Stephan de Bievre [Université de Lille, Professor, HDR]
Stefano Olla [Univ de Dauphine, Professor, until Aug 2018, HDR]
Andre de Laire [Université de Lille, Associate Professor]
Adrien Hardy [Université de Lille, Associate Professor, from Mar 2018]

PhD Student

Pierre Mennuni [Université de Lille]

Post-Doctoral Fellows

Matthias Ruf [Université Libre de Bruxelles, until Jan 2018]
Christopher Shirley [Université Libre de Bruxelles, until Aug 2018]

Visiting Scientist

Michele Triestino [Univ de Bourgogne, from Sep 2018 until Oct 2018]

Administrative Assistant

Karine Lewandowski [Inria]

2. Overall Objectives

2.1. Overall Objectives

The MEPHYSTO-POST team is a follow up of the MEPHYSTO project-team. Since the former scientific leader, Antoine Gloria, left in September 2017, the scientific objectives have been modified.

The MEPHYSTO-POST team gathers mathematicians from different communities with the same motivation: to provide a better understanding of dynamical phenomena involving particles. These phenomena are described by fundamental models arising from several fields of physics. We focus on model derivation, study of stationary states and asymptotic behaviors, as well as links between different levels of description (e.g. micro and macro models) and numerical methods to simulate such models. Applications include nonlinear optics, thermodynamics and ferromagnetism.

3. Research Program

3.1. Time asymptotics: Stationary states, solitons, and stability issues

The team investigates existence of solitons and their link with the global dynamical behavior for nonlocal problems such as that of the Gross–Pitaevskii (GP) equation which arises in models of dipolar gases. These models, in general, also introduce nonzero boundary conditions which constitute an additional theoretical and numerical challenge. Numerous results are proved for local problems, and numerical simulations allow to verify and illustrate them, as well as making a link with physics. However, most fundamental questions are still open at the moment for nonlocal problems.

The nonlinear Schrödinger (NLS) equation finds applications in numerous fields of physics. We concentrate, in a continued collaboration with our colleagues from the physics department (PhLAM) of the Université de Lille (UdL), in the framework of the Laboratoire d'Excellence CEMPI, on its applications in nonlinear optics and cold atom physics. Issues of orbital stability and modulational instability are central here.

Another typical example of problems that the team wishes to address concerns the LL equation, which describes the dynamics of the spin in ferromagnetic materials. This equation is a fundamental model in the magnetic recording industry [33] and solitons in magnetic media are of particular interest as a mechanism for data storage or information transfer [34]. It is a quasilinear PDE involving a function that takes values on the unit sphere $\mathbb{S}^2$ of $\mathbb{R}^3$. Using the stereographic projection, it can be seen as a quasilinear Schrödinger equation and the questions about the solitons, their dynamics and potential blow-up of solutions evoked above are also relevant in this context. This equation is less understood than the NLS equation: even the Cauchy theory is not completely done [28], [25]. In particular, the geometry of the target sphere imposes nonvanishing boundary conditions; even in dimension one, there are kink-type solitons having different limits at $\pm\infty$.

3.2. Derivation of macroscopic laws from microscopic dynamics

The team investigates, from a microscopic viewpoint, the dynamical mechanism at play in the phenomenon of relaxation towards thermal equilibrium for large systems of interacting particles. For instance, a first step consists in giving a rigorous proof of the fact that a particle repeatedly scattering of random obstacles through a Hamiltonian scattering process will eventually reach thermal equilibrium, thereby completing previous work in this direction by the team. As a second step, similar models as the ones considered classically will be defined and analysed in the quantum mechanical setting, and more particularly in the setting of quantum optics.

Another challenging problem is to understand the interaction of large systems with the boundaries, which is responsible for most energy exchanges (forcing and dissipation), even though it is concentrated in very thin layers. The presence of boundary conditions to evolution equations sometimes lacks understanding from a physical and mathematical point of view. In order to legitimate the choice done at the macroscopic level of the mathematical definition of the boundary conditions, we investigate systems of atoms (precisely chains of oscillators) with different local microscopic defects. We apply our recent techniques to understand how anomalous (in particular fractional) diffusive systems interact with the boundaries. For instance, the powerful tool given by Wigner functions that we already used has been successfully applied to the derivation of anomalous behaviors in open systems (for instance in [67]). The next step consists in developing an extension of that tool to deal with bounded systems provided with fixed boundaries. We also intend to derive anomalous diffusion by adding long range interactions to diffusive models. There are very few rigorous results in this direction. Finally, we aim at obtaining from a microscopic description the fractional porous medium equation (FPME), a nonlinear variation of the fractional diffusion equation, involving the fractional Laplacian instead of the usual one. Its rigorous study carries out many mathematical difficulties in treating at the same time the nonlinearity and fractional diffusion.

3.3. Numerical methods: analysis and simulations

The team addresses both questions of precision and numerical cost of the schemes for the numerical integration of nonlinear evolution PDEs, such as the NLS equation. In particular, we to develop, study and implement numerical schemes with high order that are more efficient. We also to contribute to the design and analysis of schemes with appropriate qualitative properties. These properties may as well be “asymptotic preserving” properties, energy-preserving properties, or convergence to an equilibrium properties. Other numerical goals of the team include the numerical simulation of standing waves of nonlinear nonlocal GP equations. We also keep on developing numerical methods to efficiently simulate and illustrate theoretical results on instability, in particular in the context of the modulational instability in optical fibers, where we study the influence of randomness in the physical parameters of the fibers.

4. New Results

4.1. Exponential time-decay for discrete Fokker–Planck equations

G. Dujardin and his coauthors proposed and studied in [22] several discrete versions of homogeneous and inhomogeneous one-dimensional Fokker-Planck equations. They proved in particular, for these discretizations of velocity and space, the exponential convergence to the equilibrium of the solutions, for time-continuous equations as well as for time-discrete equations. Their method uses new types of discrete Poincaré inequalities for a “two-direction” discretization of the derivative in velocity. For the inhomogeneous problem, they adapted hypocoercive methods to the discrete level.

4.2. Energy preserving methods for nonlinear Schrödinger equations

G. Dujardin and his coauthors have revisited and extended relaxation methods for nonlinear Schrödinger equations (NLS). The classical relaxation method for NLS is an energy preserving method and a mass preserving method. Moreover, it is only linearly implicit. A first proof of the second order accuracy was achieved in [14]. Moreover, the method was extended to enable to treat noncubic nonlinearities, nonlocal nonlinearities, as well as rotation terms. The resulting methods are still energy preserving and mass preserving. Moreover, they are shown to have second order accuracy numerically. These new methods are compared with fully implicit, mass and energy preserving methods of Crank and Nicolson.

4.3. Diffusive and superdiffusive behavior in one-dimensional chains of oscillators

In order to understand abnormally diffusive phenomena which are physically observed in nanotube technologies, one mathematical approach consists in starting from deterministic system of Newtonian particles, and then perturb this system with a stochastic component which provides enough ergodicity to the dynamics. It is already well known that these stochastic chains model correctly the behavior of the conductivity [24]. In [1], [2] (published in Communications in Mathematical Physics) M. Simon with her coauthors C. Bernardin, P. Gonçalves, M. Jara, T. Komorowski, S. Olla and M. Sasada have observed both behaviors, normal and anomalous diffusion, in the context of low dimensional asymmetric systems. They manage to describe the microscopic phenomena at play which are responsible for each one of these phenomena, and they go beyond the predictions that have recently been done in [31], [32]. Moreover, in [2], the authors manage to treat rigorously, for the first time, the case of an anharmonic potential: more precisely, they consider a small quartic anharmonicity and show that the result obtained in the harmonic (linear) case persists up to some small critical value of the nonlinear perturbation.

4.4. Microscopic description of moving interfaces

A large variety of models has been introduced to describe the evolution of a multiphase medium, *e.g.* the joint evolution of liquid and solid phases. These complex physical phenomena often feature absorbing phase transitions. For instance, the porous medium equation (PME)

$$\partial_t \rho = \mathrm{div}(\rho^{m-1}\, \nabla \rho), \tag{3}$$

where $m > 1$ is a constant and div and ∇ are the divergence and gradient operators in $\mathbb{R}^d$, describes the evolution of the density $\rho : \mathbb{R}^d \times \mathbb{R}_+ \to [0,1]$ of an ideal gas flowing in a homogeneous medium. It is known that, starting from an initial density ρ_0 with compact support, the solution $\rho(x,t)$ is nonnegative and has compact support in the space variable for each positive t. Thus there are interfaces separating the regions where ρ is positive from those where it is zero.

In one submitted paper in collaboration with O. Blondel, C. Cancès, and M. Sasada, we have derived the PME (1) from a degenerate and conservative dynamics in [15], for any integer $m > 1$. More precisely we improved the results previously obtained in [26], since we allow the solutions to feature moving interfaces, namely the initial condition may vanish. This moving boundary was not well apprehended at the microscopic level. Its rigorous definition is indeed very delicate, and its behavior (such that its speed, or fluctuation), as well as the relationship between the microscopic and macroscopic boundaries, are challenging questions that we aim to tackle in a near future.

When $m < 1$, equation (1) is called fast diffusion equation. In a recent collaborative work (submitted) with O. Blondel, C. Erignoux and M. Sasada [16], we derive such a fast diffusion equation in dimension one from an interacting particle system belonging to the class of conserved lattice gases with active-absorbing phase transition [30]. The microscopic dynamics is very constrained: in a few words, a particle can jump to the right (resp. left) empty neighboring site if and only if it has a particle to its left (resp. right) neighboring site. This model is really complex: the state space is divided into transient states, absorbing states and ergodic states. Depending on the initial number of particles, the transient good configurations will lead to the ergodic component and the transient bad configurations will be absorbed to an inactive state. Because of the jump constraint, there are two distinct regimes for the macroscopic behavior. Either the macroscopic density is larger than $\frac{1}{2}$, in which case the system behaves diffusively, or the density is lower than $\frac{1}{2}$, in which case the system freezes rapidly.

The interfaces between these two phases propagate as particles from the supercritical phase ($\rho > \frac{1}{2}$) diffuse towards the subcritical phase ($\rho < \frac{1}{2}$). We expect that the macroscopic density profile evolves under the diffusive scaling according to the Stefan problem

$$\partial_t \rho = \Delta\left(G(\rho)\right) \qquad \text{where } G(\rho) = \tfrac{2\rho-1}{\rho}\, \mathbf{1}_{\rho > \frac{1}{2}}. \tag{4}$$

The microscopic derivation of such Stefan problems is a well known difficult problem, only partially solved [27], [29]. In [16] we treat the liquid part of the problem (*i.e.* when the initial profiles ρ_0 are uniformly larger than the critical density $\frac{1}{2}$) and we provide a refined estimation of the time needed by the system to enter into the ergodic state. Then, we show that the macroscopic density profile evolves under the diffusive time scaling according to (1) with $m = -1$. The extension to more general initial profiles is our next goal.

4.5. Stability analysis of a Vlasov-Wave system

S. De Bièvre and his co-authors introduced and studied a kinetic equation of the Vlasov-Wave type, which arises in the description of the behavior of a large number of particles interacting weakly with an environment, composed of an infinite collection of local vibrational degrees of freedom, modeled by wave equations. They use variational techniques to establish the existence of large families of stationary states for this system, and analyze their stability [8].

4.6. Orbital stability in the presence of symmetries

With S. Rota Nodari, S. De Bièvre considered the orbital stability of relative equilibria of Hamiltonian dynamical systems on Banach spaces, in the presence of a multi-dimensional invariance group for the dynamics [9]. They proved a persistence result for such relative equilibria, presented a generalization of the Vakhitov-Kolokolov slope condition to this higher dimensional setting, and showed how it allows to prove the local coercivity of the Lyapunov function, which in turn implies orbital stability. The method was applied to study the orbital stability of relative equilibria of nonlinear Schrödinger and Manakov equations. It extends and clarifies the approach of Grillakis-Shatah-Strauss.

4.7. Measuring nonclassicality of bosonic field quantum state

S. De Bièvre and his collaborators introduced a new distance-based measure for the nonclassicality of the states of a bosonic field, which outperforms the existing such measures in several ways [17]. They defined for that purpose the operator ordering sensitivity of the state which evaluates the sensitivity to operator ordering of the Renyi entropy of its quasi-probabilities and which measures the oscillations in its Wigner function. Through a sharp control on the operator ordering sensitivity of classical states they obtained a precise geometric image of their location in the density matrix space allowing them to introduce a distance-based measure of nonclassicality. They analyze the link between this nonclassicality measure and a recently introduced quantum macroscopicity measure, showing how the two notions are distinct.

4.8. The Cauchy problem for the Landau–Lifshitz–Gilbert equation in BMO and self-similar solutions

A. de Laire and S. Gutierrez established in [19] a global well-posedness result for the Landau–Lifshitz equation with Gilbert damping, provided that the BMO semi-norm of the initial data is small. As a consequence, they deduced the existence of self-similar solutions in any dimension. Moreover, in the one-dimensional case, they characterized the self-similar solutions when the initial data is given by some ($\S^2$-valued) step function and established their stability. They also showed the existence of multiple solutions if the damping is strong enough.

4.9. The Sine–Gordon regime of the Landau–Lifshitz equation with a strong easy-plane anisotropy

It is well-known that the dynamics of biaxial ferromagnets with a strong easy-plane anisotropy is essentially governed by the Sine-Gordon equation. A. de Laire and P. Gravejat provided in [10] a rigorous justification to this observation. More precisely, they showed the convergence of the solutions to the Landau-Lifshitz equation for biaxial ferromagnets towards the solutions to the Sine-Gordon equation in the regime of a strong easy-plane anisotropy. This result holds for solutions to the Landau–Lifshitz equation in high order Sobolev spaces. They also provided an alternative proof for local well-posedness in this setting by introducing high order energy quantities with better symmetrization properties. Then they derived the convergence from the consistency of the Landau–Lifshitz equation with the Sine-Gordon equation by using well-tailored energy estimates. As a by-product, they also obtained a further derivation of the free wave regime of the Landau–Lifshitz equation.

4.10. Mutual information of wireless channels and block-Jacobi ergodic operators

In telecommunication models the quality of the transferred data is assessed through the entropy of the channel, a theoretical quantity that is usually not computable in practice. W. Hachem, A. Hardy and S. Shamai prove in [23] that one can relate this quantity for a large class of models involving several antennas (MIMO) to the equilibrium measure of a matrix valued Markov chain associated with the model, and so does its asymptotic behavior when the signal-noise-ratio parameter becomes large. By means of ergodicity results, this yields estimates for these quantities that are implementable faster than the naive estimators.

4.11. DLR equations and rigidity for the Sine-beta process

The Sine-beta process is a universal object appearing in the study of large Hermitian random matrices and statistical systems in a logarithmic interaction, such as low dimensional Coulomb gases. However, the only description available yet relied on a rather complicated and non-physical system of coupled stochastic differential equations. In [21], D. Dereudre, A. Hardy, T. Leblé and M. Maïda obtain a statistical physic interpretation of the Sine-beta process as probability measure on infinite configurations of points described by means of the DLR formalism. This allows to obtain more information on the Sine-beta process: for instance, it is rigid, it is tolerant, and the number of particles in a compact box has gaussian fluctuations as the box becomes large.

4.12. Time-frequency transforms of white noises and Gaussian analytic functions

In signal processing, an important challenge is to be able to separate signals from ambient noises. In time-frequency analysis, this problem reduces to identify what is the spectrogram of a white noise to derive statistical tests in order to decide if some partial signal is noise or not. P. Fandrin recently put forward that the understanding of the zeros of the spectrograms would be already an important step by analyticity of the spectrograms. R. Bardenet and A. Hardy observed in [13] that there is a canonical way to identify the zeros of the usual white noise transforms associated to classical spectrograms and zeroes of Gaussian analytic functions associated with classical orthogonal polynomials in the background. In particular the zeros satisfy some invariance properties leading to computable correlation functions. In specific cases, one can identify some transforms whose zeros form a determinantal point process, in which case all the statistics of interests can be computed explicitly and this allows an exact numerical treatment.

4.13. Energy of the Coulomb gas on the sphere at low temperature

In relation to the 7th Smale problem, which is about finding polynomial time algorithm to produce well spread configuration of points on the sphere in a quantified manner, C. Beltran and A. Hardy proved in [4] that the Coulomb gas on the sphere at a temperature proportional to the inverse number of points in a configuration reaches the numerical precision required by this problem. We however did not discuss yet the algorithmic procedure, which is currently in investigation by A. Hardy and M. Simon.

4.14. Polynomial ensembles and recurrence coefficients

Determinantal point processes can be of important use in applications as soon as one is interested in producing configurations of well spread points on an arbitrary space. A class of determinantal point processes on the real line that has been extensively studied recently are the so-called polynomial ensembles. A. Hardy gathered in [11] several results concerning these models in relation to the recurrent coefficients associated with the orthogonal polynomials hidden in the background.

4.15. Concentration for Coulomb gases and Coulomb transport inequalities

The convergence of the Coulomb gas, which is a statistical gas of charged particles in an electrostatic interaction, towards its limiting distribution as the number of particles goes to infinity is a result which is part of the folklore of potential theory. The speed at which this convergence arise, which can be assessed through concentration of measure estimates in, say, the Wasserstein-Kantorovich metric, are however new results obtained by D. Chafaï, A. Hardy and M. Maïda in [7]. One of the main ingredient was to develop transport inequalities associated with the Coulomb interaction.

5. Partnerships and Cooperations

5.1. National Initiatives

5.1.1. ANR

A. de Laire is a member of the ANR ODA project.

Title: Dispersive and random waves.

ANR reference : ANR-18-CE40-0020-01.

Coordinator: Nikolay Tzvetkov, Université de Cergy-Pontoise.

A. Hardy is a member of the ANR BoB project.

Title: Inférence bayésienne à ressources limitées - données massives et modèles coûteux.

Programme ANR: (DS0705) 2016.

ANR reference: ANR-16-CE23-0003.

Coordinator: R. Bardenet, CNRS & Université de Lille.

Duration: October 2016 - October 2020.

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: C. Bernardin, Université de Nice.

Duration: October 2014 - October 2019.

5.2. European Initiatives

M. Simon is a collaborator of the ERC Starting Grant HyLEF project.

Title: Hydrodynamic Limits and Equilibrium Fluctuations: universality from stochastic systems

Duration: May 2017 - April 2022

Coordinator: P. Gonçalves, Instituto Superior Técnico, Lisbon.

5.3. International Research Visitors

5.3.1. Visits to International Teams

5.3.1.1. Research Stays Abroad

S. De Bièvre spent two months at the Centre de Recherche Mathématiques in Montréal as Simons Professor. M. Simon has been invited as Junior Scientific Leader of the Simons Semester "PDEs/SPDEs and Functional Inequalities" at IMPAN in Warsaw, Poland, for one month.

6. Dissemination

6.1. Promoting Scientific Activities

6.1.1. Scientific Events Organisation

6.1.1.1. Member of the Organizing Committees

A. Hardy co-organized the "Semaine d'Etude Math-Entreprise Hauts de France 2018" (Lille).

6.1.2. Journal

6.1.2.1. Reviewer - Reviewing Activities

- S. De Bièvre served as reviewer for J. Math. Phys., Ann. Institut H. Poincaré, J. Stat. Phys. in 2018.
- G. Dujardin served as reviewer for APNUM and Numer. Math. in 2018.
- A. Hardy served as reviewer for Communications in Pure and Applied Mathematics and Annals of Applied Probability in 2018.
- M. Simon was reviewer for Markov Processes and Related Fields, and Annales de L'I.H.P. Probabilités et Statistiques in 2018.

6.1.3. Invited Talks

A. Hardy was invited to give several talks in 2018, including:

- (May 2018) Workshop "random matrices and their applications", Kyoto university (Japan)
- (April 2018) Groupe de travail "Probas du vendredi" de Jussieu, Paris
- A. Hardy was invited at a "réunion interne de l'Académie des Science" entitled "Le renouveau des processus ponctuels déterminantaux, des fermions à la statistique appliquée".

M. Simon was invited to give several talks in 2018, including:

- (April 2018) Workshop of the Simons Semester "PDE/SPDE-s, Functional Inequalities", Banach Center, Poznan (Poland)
- (August 2018) Journées "Modélisation Aléatoire et Statistique" of the "Société de Mathématiques Appliquées et Industrielles", Dijon (France)
- (November 2018) Weekly Probability Seminar at University of Bath (England)
- (December 2018) Weekly Probability Seminar in Lyon (France).

6.1.4. Research Administration

G. Dujardin is a member of Inria Evaluation Committee.

6.2. Teaching - Supervision - Juries

6.2.1. Teaching

Licence: G. Dujardin, "Calcul Différentiel et Intégral", 30h, L2, Université Libre de Bruxelles, Belgique.

Master: G. Dujardin, "Analyse Fonctionelle", 30h, M1, Université Libre de Bruxelles, Belgique.

Master: G. Dujardin, "Vortex dans les condensats de Bose–Einstein en rotation", 20h, M2, Université de Lille, France.

6.2.2. Supervision

HdR: Guillaume Dujardin, Contribution à l'analyse numérique de problèmes d'évolution : comportements asymptotiques et applications à l'équation de Schrödinger, Université de Lille, November 12th 2018 [3].

6.2.3. Juries

G. Dujardin and M. Simon participated in the jury of the "Agrégation externe de mathématiques" in 2018.
G. Dujardin took part in the hiring committees of Junior Scientists for Inria Paris, Inria Saclay and in the final admission committee in 2018.
M. Simon was member of the jury of the PhD thesis of J. Roussel which was defended in November 2018 at École des Ponts (Marne-la-Vallée, France) and is entitled *Theoretical and numerical analysis of non-reversible dynamics in computational statistical physics*.

6.3. Popularization

6.3.1. Interventions

- M. Simon participated in the local program “Chercheurs itinérants”, and gave several lectures directed to high-school students.

7. Bibliography

Major publications by the team in recent years

[1] C. Bernardin, P. Gonçalves, M. Jara, M. Simon.*Interpolation process between standard diffusion and fractional diffusion*, August 2017, to appear in AIHP B, https://hal.archives-ouvertes.fr/hal-01348503

[2] C. ´. Bernardin, P. Gonçalves, M. Jara, M. Simon.*Nonlinear Perturbation of a Noisy Hamiltonian Lattice Field Model: Universality Persistence*, August 2017, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01491433

Publications of the year

Doctoral Dissertations and Habilitation Theses

[3] G. Dujardin.*Contribution à l’analyse numérique de problèmes d’évolution : comportements asymptotiques et applications à l’équation de Schrödinger*, Universite de Lille, November 2018, Habilitation à diriger des recherches, https://hal.archives-ouvertes.fr/tel-01950160

Articles in International Peer-Reviewed Journal

[4] C. Beltrán, A. Hardy.*Energy of the Coulomb Gas on the Sphere at Low Temperature*, in "Archive for Rational Mechanics and Analysis", October 2018 [*DOI :* 10.1007/s00205-018-1316-3], https://hal.archives-ouvertes.fr/hal-01890125

[5] C. Bernardin, P. Gonçalves, M. Jara, M. Simon.*Interpolation process between standard diffusion and fractional diffusion*, in "Annales de l’Institut Henri Poincaré (B) Probabilités et Statistiques", 2018, https://hal.archives-ouvertes.fr/hal-01348503

[6] C. Bernardin, P. Gonçalves, M. Jara, M. Simon.*Nonlinear Perturbation of a Noisy Hamiltonian Lattice Field Model: Universality Persistence*, in "Communications in Mathematical Physics", 2018, https://hal.archives-ouvertes.fr/hal-01491433

[7] D. Chafai, A. Hardy, M. Maïda.*Concentration for Coulomb gases and Coulomb transport inequalities*, in "Journal of Functional Analysis", September 2018, vol. 275, n° 16, p. 1447-1483, https://arxiv.org/abs/1610.00980 [*DOI :* 10.1016/j.jfa.2018.06.004], https://hal.archives-ouvertes.fr/hal-01374624

[8] S. De Bièvre, T. Goudon, A. Vavasseur.*Stability analysis of a Vlasov-Wave system describing particles interacting with their environment*, in "Journal of Differential Equations", June 2018, vol. 264, n° 12, p. 7069-7093, https://hal.inria.fr/hal-01581676

[9] S. De Bièvre, S. Rota Nodari.*Orbital stability via the energy-momentum method: the case of higher dimensional symmetry groups*, in "Archive for Rational Mechanics and Analysis", 2018, https://arxiv.org/abs/1605.02523 [*DOI :* 10.1007/s00205-018-1278-5], https://hal.archives-ouvertes.fr/hal-01312534

[10] A. DE LAIRE, P. GRAVEJAT.*The Sine-Gordon regime of the Landau-Lifshitz equation with a strong easy-plane anisotropy*, in "Annales de l'Institut Henri Poincare (C) Analyse Non Lineaire", November 2018, vol. 35, n^{o} 7, p. 1885-1945, https://hal.archives-ouvertes.fr/hal-01518483

[11] A. HARDY.*Polynomial Ensembles and Recurrence Coefficients*, in "Constructive Approximation", August 2018, vol. 48, n^{o} 1, p. 137 - 162 [*DOI :* 10.1007/s00365-017-9413-3], https://hal.archives-ouvertes.fr/hal-01890050

[12] T. KOMOROWSKI, S. OLLA, M. SIMON.*Macroscopic evolution of mechanical and thermal energy in a harmonic chain with random flip of velocities*, in "Kinetic and Related Models ", 2018, vol. 11, n^{o} 3, p. 615-645, https://hal.archives-ouvertes.fr/hal-01358979

Other Publications

[13] R. BARDENET, A. HARDY.*Time-frequency transforms of white noises and Gaussian analytic functions*, August 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01855678

[14] C. BESSE, S. DESCOMBES, G. DUJARDIN, I. LACROIX-VIOLET.*Energy preserving methods for nonlinear schrodinger equations*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01951527

[15] O. BLONDEL, C. CANCÈS, M. SASADA, M. SIMON.*Convergence of a Degenerate Microscopic Dynamics to the Porous Medium Equation*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01710628

[16] O. BLONDEL, C. ERIGNOUX, M. SASADA, M. SIMON.*Hydrodynamic limit for an activated exclusion process*, May 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01798051

[17] S. DE BIÈVRE, D. B. HOROSHKO, G. PATERA, M. I. KOLOBOV.*Measuring nonclassicality of bosonic field quantum states via operator ordering sensitivity*, August 2018, Preprint, https://hal.archives-ouvertes.fr/hal-01865597

[18] A. DE LAIRE, P. GRAVEJAT.*The cubic Schrödinger regime of the Landau-Lifshitz equation with a strong easy-axis anisotropy*, December 2018, https://arxiv.org/abs/1812.05854 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01954762

[19] A. DE LAIRE, S. GUTIÉRREZ.*The Cauchy problem for the Landau-Lifshitz-Gilbert equation in BMO and self-similar solutions*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01948679

[20] A. DE LAIRE, P. MENNUNI.*Traveling waves for some nonlocal 1D Gross-Pitaevskii equations with nonzero conditions at infinity*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01962779

[21] D. DEREUDRE, A. HARDY, T. LEBLÉ, M. MAÏDA.*DLR equations and rigidity for the Sine-beta process*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01954367

[22] G. DUJARDIN, F. HÉRAU, P. LAFITTE-GODILLON.*Coercivity, hypocoercivity, exponential time decay and simulations for discrete Fokker- Planck equations*, February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01702545

[23] W. HACHEM, A. HARDY, S. SHAMAI.*Mutual Information of Wireless Channels and Block-Jacobi Ergodic Operators*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01954454

References in notes

[24] G. BASILE, C. BERNARDIN, S. OLLA.*Thermal conductivity for a momentum conservative model*, in "Comm. Math. Phys.", 2009, vol. 287, n^{o} 1, p. 67–98, http://dx.doi.org/10.1007/s00220-008-0662-7

[25] I. BEJENARU, A. D. IONESCU, C. E. KENIG, D. TATARU.*Global Schrödinger maps in dimensions $d \geq 2$: small data in the critical Sobolev spaces*, in "Annals of Mathematics", 2011, p. 1443–1506

[26] P. GONÇALVES, C. LANDIM, C. TONINELLI.*Hydrodynamic limit for a particle system with degenerate rates*, in "Ann. Inst. Henri Poincaré Probab. Stat.", 2009, vol. 45, n^{o} 4, p. 887–909, https://doi.org/10.1214/09-AIHP210

[27] J. GRAVNER, J. QUASTEL.*Internal DLA and the Stefan problem*, in "Ann. Probab.", 2000, vol. 28, n^{o} 4, p. 1528–1562, https://doi.org/10.1214/aop/1019160497

[28] R. L. JERRARD, D. SMETS.*On Schrödinger maps from T^1 to S^2*, in "Ann. Sci. ENS", 2012, vol. 45, p. 637-680

[29] C. LANDIM, G. VALLE.*A microscopic model for Stefan's melting and freezing problem*, in "Ann. Probab.", 2006, vol. 34, n^{o} 2, p. 779–803, https://doi.org/10.1214/009117905000000701

[30] M. ROSSI, R. PASTOR-SATORRAS, A. VESPIGNANI.*Universality Class of Absorbing Phase Transitions with a Conserved Field*, in "Phys. Rev. Lett.", 2000, vol. 85, n^{o} 1803

[31] H. SPOHN.*Nonlinear Fluctuating Hydrodynamics for Anharmonic Chains*, in "Journal of Statistical Physics", Mar 2014, vol. 154, n^{o} 5, p. 1191–1227, https://doi.org/10.1007/s10955-014-0933-y

[32] H. SPOHN.*The Kardar-Parisi-Zhang equation – a statistical physics perspective*, in "Arxiv preprint 1601.00499", 01 2016

[33] D. WEI.*Micromagnetics and Recording Materials*, Springer–Verlag Berlin Heidelberg, 2012, http://dx.doi.org/10.1007/978-3-642-28577-6 , https://doi.org/10.1007/978-3-642-28577-6

[34] S. ZHANG, A. A. BAKER, S. KOMINEAS, T. HESJEDAL.*Topological computation based on direct magnetic logic communication*, in "Scientific Reports", 2015, vol. 5, http://dx.doi.org/10.1038/srep15773

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

1. Team, Visitors, External Collaborators **257**
2. Overall Objectives **258**
2.1. Context 258
2.2. Goals 258
3. Research Program **259**
3.1. Research axis 1: Unsupervised learning 259
3.2. Research axis 2: Performance assessment 259
3.3. Research axis 3: Functional data 259
3.4. Research axis 4: Applications motivating research 259
4. Application Domains **260**
4.1. Economic World 260
4.2. Biology 260
5. Highlights of the Year **260**
6. New Software and Platforms **260**
6.1. MixtComp 260
6.2. BlockCluster 260
6.3. CloHe 261
6.4. PACBayesianNMF 261
6.5. pycobra 261
6.6. STK++ 262
6.7. rtkore 262
6.8. MixAll 262
6.9. simerge 263
6.10. Platforms 263
7. New Results **263**
7.1. Axis 1: Data Units Selection in Statistics 263
7.2. Axis 1: Model-Based Co-clustering for Ordinal Data 263
7.3. Axis 1: Model-Based Co-clustering for Ordinal Data of different dimensions 264
7.4. Axis 1: Model-based co-clustering for mixed type data 264
7.5. Axis 1: Model-Based Co-clustering with Co-variables 264
7.6. Axis 1: Relaxing the Identically Distributed Assumption in Gaussian Co-Clustering for High Dimensional Data 265
7.7. Axis 1: Gaussian-based visualization of Gaussian and non-Gaussian model-based clustering 265
7.8. Axis 1: A targeted multi-partitions clustering 265
7.9. Axis 1: Co-clustering: A versatile way to perform clustering in high dimension 265
7.10. Axis 1: Dealing with missing data in model-based clustering through a MNAR model 266
7.11. Axis 1: Self Organizing Coclustering for textual data synthesis 266
7.12. Axis 1: Linking canonical and spectral clustering 267
7.13. Axis 1: Multiple partition clustering 267
7.14. Axis 2: Change-point detection by means of reproducing kernels 267
7.15. Axis 2: New efficient algorithms for multiple change-point detection with kernels 267
7.16. Axis 2: Multi-Layer Group-Lasso 268
7.17. Axis 2: Pseudo-Bayesian Learning with Kernel Fourier Transform as Prior 268
7.18. Axis 2: Decentralized learning with budgeted network load using Gaussian copulas and classifier ensembles 268
7.19. Axis 2: Sequential Learning of Principal Curves: Summarizing Data Streams on the Fly 269
7.20. Axis 2: A Quasi-Bayesian Perspective to Online Clustering 269
7.21. Axis 2: Pycobra: A Python Toolbox for Ensemble Learning and Visualisation 269
7.22. Axis 2: Simpler PAC-Bayesian bounds for hostile data 269

7.23. Axis 2: PAC-Bayesian high dimensional bipartite ranking 270
7.24. Axis 2: Multiview Boosting by Controlling the Diversity and the Accuracy of View-specific Voters 270
7.25. Axis 3: Clustering spatial functional data 270
7.26. Axis 3: Categorical functional data analysis 270
7.27. Axis 4: Real-time Audio Sources Classification 271
7.28. Axis 4: Matching of descriptors evolving over time 271
7.29. Axis 4: Supervised multivariate discretization and levels merging for logistic regression 271
7.30. Axis 4: MASSICCC Platform for SaaS Software Availability 272
7.31. Axis 4: ClinMine: Optimizing the Management of Patients in Hospital 272
7.32. Projection Under Pairwise Control 272
8. Bilateral Contracts and Grants with Industry 272
8.1. Bilateral Contracts: SEMENCES DE FRANCE 272
8.2. Bilateral Contracts: Arcelor-Mittal 273
8.3. Bilateral Contracts: Alstom 273
8.4. Bilateral Contracts: Decathlon 273
9. Partnerships and Cooperations 273
9.1. Regional Initiatives 273
9.1.1. Bilille partnership 273
9.1.2. Bilille collaborations 274
9.2. National Initiatives 274
9.2.1. Programme of Investments for the Future (PIA) 274
9.2.2. RHU PreciNASH 274
9.2.3. INS2I-CNRS project PEPS JCJC 2018 “PaRaFF” 275
9.2.4. ANR 275
9.2.4.1. ANR APRIORI 275
9.2.4.2. ANR BEAGLE 275
9.2.4.3. ANR SMILE 275
9.2.4.4. ANR ClinMine 276
9.2.4.5. ANR TheraSCUD2022 276
9.2.5. Working groups 276
9.2.6. Other initiatives 276
9.3. European Initiatives 277
9.3.1. FP7 & H2020 Projects 277
9.3.2. Collaborations with Major European Organizations 277
9.4. International Initiatives 277
9.4.1. Inria International Labs 277
9.4.2. Participation in Other International Programs 278
9.5. International Research Visitors 278
10. Dissemination 279
10.1. Promoting Scientific Activities 279
10.1.1. Scientific Events Organisation 279
10.1.2. Scientific Events Selection 279
10.1.2.1. Chair of Conference Program Committees 279
10.1.2.2. Member of the Conference Program Committees 279
10.1.2.3. Reviewer 279
10.1.3. Journal 279
10.1.4. Invited Talks 280
10.1.5. Leadership within the Scientific Community 280
10.1.6. Scientific Expertise 280
10.1.7. Research Administration 280

10.2. Teaching - Supervision - Juries 281
10.2.1. Teaching 281
10.2.2. Supervision 282
10.2.3. Juries 282
10.3. Popularization 282
10.3.1. Internal or external Inria responsibilities 282
10.3.2. Internal action 282
11. Bibliography **282**

Project-Team MODAL

Creation of the Team: 2010 September 01, updated into Project-Team: 2012 January 01

Keywords:

Computer Science and Digital Science:

A3.1.4. - Uncertain data
A3.2.3. - Inference
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.5. - Bayesian methods
A3.4.7. - Kernel methods
A5.2. - Data visualization
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.3.3. - Data processing
A9.2. - Machine learning

Other Research Topics and Application Domains:

B2.2.3. - Cancer
B9.5.6. - Data science
B9.6.3. - Economy, Finance
B9.6.5. - Sociology

1. Team, Visitors, External Collaborators

Research Scientists

Christophe Biernacki [Team leader, Inria, Senior Researcher, HDR]
Pascal Germain [Inria, Researcher]
Benjamin Guedj [Inria, Researcher]
Hemant Tyagi [Inria, Researcher, from Dec 2018]

Faculty Members

Alain Celisse [Univ de Lille, Associate Professor, HDR]
Serge Iovleff [Univ de Lille, Associate Professor]
Guillemette Marot [Univ de Lille, Associate Professor]
Cristian Preda [Univ de Lille, Professor, HDR]
Vincent Vandewalle [Univ de Lille, Associate Professor]
Sophie Dabo [Univ de Lille, Associate Professor]

External Collaborators

Julien Jacques [Univ Lumière, HDR]
Philippe Heinrich [Univ de Lille]

Technical Staff

Vincent Kubicki [Inria, from Oct 2018]
Bhargav Srinivasa Desikan [Inria, until Aug 2018]

PhD Students
Yaroslav Averyanov [Inria]
Maxime Baelde [A-Volute]
Anne-Lise Bedenel [MeilleureAssurance]
Adrien Ehrhardt [CA CF]
Arthur Leroy [Univ René Descartes]
Le Li [iAdvize & Univ. dAngers, until November 2018]

Post-Doctoral Fellow
Fabien Laporte [Ecole polytechnique, from Jul 2018]

Visiting Scientist
Michael Gallaugher [McMaster University, from Mar 2018 until May 2018]

Administrative Assistant
Anne Rejl [Inria]

2. Overall Objectives

2.1. Context

In several respects, modern society has strengthened the need for statistical analysis, even if other related names are sometimes preferably used depending on methods, communities and applications, as data analysis, machine learning or artificial intelligence. The genesis comes from the easier availability of data thanks to technological breakthroughs (storage, transfer, computing), and are now so widespread that they are no longer limited to large human organizations. The more or less conscious goal of such data availability is the expectation to improving the quality of “since the dawn of time” statistical stories which are namely discovering new knowledge or doing better predictions. These both central tasks can be referred respectively as unsupervised learning or supervised learning, even if it is not limited to them or other names exist depending on communities. Somewhere, it pursues the following hope: “more data for better and more results”.

However, today’s data are more and more complex. They gather mixed type features (for instance continuous data mixed with categorical data), missing or partially missing (like intervals) items and numerous variables (high dimensional situation). As a consequence, the target “better and more results” of the previous adage (both words are important: “better” and also “more”) could not be reached through a somewhat “handwork” way, but should inevitably rely on some theoretical formalization and guarantee. Indeed, data can be so numerous and so complex (data can live in quite abstract spaces) that the “empirical” statistician is quickly outdated. However, data being subject by nature to randomness, the probabilistic framework is a very sensible theoretical environment to serve as a general guide for modern statistical analysis.

2.2. Goals

Modal is a project-team working on today’s complex data sets (mixed data, missing data, high-dimensional data), for classical statistical targets (unsupervised learning, supervised learning, regression, ...) with approaches relying on the probabilistic framework. This latter can be tackled through both model-based methods (as mixture models for a generic tool) and model-free methods (as probabilistic bounds on empirical quantities). Furthermore, Modal is connected to the real world by applications, typically with biological ones (some members have this skill) but many other ones are also considered since the application coverage of the Modal methodology is very large. It is also important to note that, in return, applications are often real opportunities for initiating academic questioning for the statistician (case of the Bilille platform and some bilateral contracts of the team).

From the academic communities point of view, Modal can be seen as belonging simultaneously to both the statistical learning and machine learning ones, as attested by its publications. Somewhere it is the opportunity to make a bridge between these two stochastic communities around a common but large probabilistic framework.

3. Research Program

3.1. Research axis 1: Unsupervised learning

Scientific locks related to unsupervised learning are numerous, concerning the clustering outcome validity, the ability to manage different kinds of data, the missing data questioning, the dimensionality of the data set, ... Many of them are addressed by the team, leading to publication achievements, often with a specific package delivery (sometimes upgraded as a software or even as a platform grouping several software). Because of the variety of the scope, it involves nearly all the permanent team members, often with PhD students and some engineers. The related works are always embedded inside a probabilistic framework, typically model-based approaches but also model-free ones like PAC-Bayes (PAC stands for Probably Approximately Correct), because such a mathematical environment offers both a well-posed problem and a rigorous answer.

3.2. Research axis 2: Performance assessment

One main concern of the Modal team is to provide theoretical justifications on the procedures which are designed. Such guarantees are important to avoid misleading conclusions resulting from any unsuitable use. The main ingredient in proving these guarantees is the use of the PAC framework, leading to finite-sample concentration inequalities. More precisely, contributions to PAC learning rely on the classical empirical process theory and the PAC-Bayesian theory. The Modal team exploits these non-asymptotic tools to analyze the performance of iterative algorithms (such as gradient descent), cross-validation estimators, online change-point detection procedures, ranking algorithms, matrix factorization techniques and clustering methods, for instance. The team also develops some expertise on the formal dynamic study of algorithms related to mixture models (important models used in the previous unsupervised setting), like degeneracy for EM algorithm or also label switching for Gibbs algorithm.

3.3. Research axis 3: Functional data

Mainly due to technological advances, functional data are more and more widespread in many application domains. Functional data analysis (FDA) is concerned with the modeling of data, such as curves, shapes, images or a more complex mathematical object, though as smooth realizations of a stochastic process (an infinite dimensional data object valued in a space of eventually infinite dimension; space of squared integrable functions, ...). Time series are an emblematic example even if it should not be limited to them (spectral data, spatial data, ...). Basically, FDA considers that data correspond to realizations of stochastic processes, usually assumed to be in a metric, semi-metric, Hilbert or Banach space. One may consider, functional independent or dependent (in time or space) data objects of different types (qualitative, quantitative, ordinal, multivariate, time-dependent, spatial-dependent, ...). The last decade saw a dynamic literature on parametric or non-parametric FDA approaches for different types of data and applications to various domains, such as principal component analysis, clustering, regression and prediction.

3.4. Research axis 4: Applications motivating research

The fourth axis consists in translating real application issues into statistical problems raising new (academic) challenges for models developed in Modal team. Cifre Phds in industry and interdisciplinary projects with research teams in Health and Biology are at the core of this objective. The main originality of this objective lies in the use of statistics with complex data, including in particular ultra-high dimension problems. We focus on real applications which cannot be solved by classical data analysis.

4. Application Domains

4.1. Economic World

The Modal team applies it research to the economic world through CIFRE Phd supervision such as CACF (credit scoring), A-Volute (expert in 3D sound), Meilleur Taux (insurance comparator), ...It also has many contracts with companies such as Decathlon (world leader in sports equipment), Arcelor-Mittal (steel industry) or Alstom (integrated transport systems).

4.2. Biology

The second main application domain of the team is the biology. Members of the team are involved in the supervision and scientific animation of the bilille platform, the bioinformatics and bioanalysis platform of Lille.

5. Highlights of the Year

5.1. Highlights of the Year

- Hemant Tyagi has been recruited as CR in the team.
- Three new ANR funded (one began in 2018, two will start in 2019).
- One H2020 European project funded (started in November 2018).
- One-year EIT European project called SysBooster with ApSys and Nokia.
- Creation of a startup using MODAL's technology (MixtComp software).

6. New Software and Platforms

6.1. 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). 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: Christophe Biernacki, Étienne Goffinet, Matthieu Marbac-Lourdelle, Quentin Grimonprez, Serge Iovleff and Vincent Kubicki
- Contact: Christophe Biernacki
- URL: https://modal-research.lille.inria.fr/BigStat

6.2. 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: Christophe Biernacki, Gilles Celeux, Parmeet Bhatia, Serge Iovleff, Vincent Brault and Vincent Kubicki
- Partner: Université de Technologie de Compiègne
- Contact: Serge Iovleff
- URL: http://cran.r-project.org/web/packages/blockcluster/index.html

6.3. CloHe

Clustering of Mixed data
KEYWORDS: Classification - Clustering - Missing data
FUNCTIONAL DESCRIPTION: Software of classification for mixed data with missing values with application to multispectral satellite image time-series

- Partners: CNRS - INRA
- Contact: Serge Iovleff
- URL: https://modal.lille.inria.fr/CloHe/

6.4. PACBayesianNMF

KEYWORDS: Statistics - Machine learning
FUNCTIONAL DESCRIPTION: Implementing NMF with a PAC-Bayesian approach relying upon block gradient descent

- Participants: Benjamin Guedj and Astha Gupta
- Contact: Benjamin Guedj
- URL: https://github.com/astha736/PACbayesianNMF

6.5. pycobra

KEYWORDS: Statistics - Data visualization - Machine learning
SCIENTIFIC DESCRIPTION: pycobra is a python library for ensemble learning, which serves as a toolkit for regression, classification, and visualisation. It is scikit-learn compatible and fits into the existing scikit-learn ecosystem.

pycobra offers a python implementation of the COBRA algorithm introduced by Biau et al. (2016) for regression.

Another algorithm implemented is the EWA (Exponentially Weighted Aggregate) aggregation technique (among several other references, you can check the paper by Dalalyan and Tsybakov (2007).

Apart from these two regression aggregation algorithms, pycobra implements a version of COBRA for classification. This procedure has been introduced by Mojirsheibani (1999).

pycobra also offers various visualisation and diagnostic methods built on top of matplotlib which lets the user analyse and compare different regression machines with COBRA. The Visualisation class also lets you use some of the tools (such as Voronoi Tesselations) on other visualisation problems, such as clustering.

- Participants: Bhargav Srinivasa Desikan and Benjamin Guedj
- Contact: Benjamin Guedj
- Publication: Pycobra: A Python Toolbox for Ensemble Learning and Visualisation
- URL: https://github.com/bhargavvader/pycobra

6.6. STK++

Statistical ToolKit
KEYWORDS: Statistics - Linear algebra - Framework - Learning - Statistical learning
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

6.7. 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

6.8. 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 (www.stkpp.org) for faster computation.

- Participant: Serge Iovleff
- Partner: Université Lille 1
- Contact: Serge Iovleff
- URL: https://cran.r-project.org/web/packages/MixAll/

6.9. simerge

Statistical Inference for the Management of Extrem Risks, Genetics and Global epidemiology
KEYWORD: Biclustering
FUNCTIONAL DESCRIPTION: Allows to perform Co-Clustering on binary (Bernoulli) and counting variables (Poisson) using co-variables.

- Partner: Inria
- Contact: Serge Iovleff

6.10. Platforms

6.10.1. MASSICCC Platform

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, the BlockCluster package has been integrated and also a particular attention to provide meaningful graphical outputs (for Mixmod, MixtComp and BlockCluster) directly in the web platform itself has led to some specific developments.

7. New Results

7.1. Axis 1: Data Units Selection in Statistics

Participant: Christophe Biernacki.

Usually, the data unit definition is fixed by the practitioner but it can happen that he/her 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, count, categorical). This work is now published in an international journal [12].

An extension of this work has been also presented to an international workshop.The idea is to use the data units principle as a way for (co-)clustering model enlargement.

It is a joint work with Alexandre Lourme from University of Bordeaux.

7.2. Axis 1: Model-Based Co-clustering for Ordinal Data

Participant: Christophe Biernacki.

A model-based co-clustering algorithm for ordinal data is presented. This algorithm relies on the latent block model embedding a probability distribution specific to ordinal data (the so-called BOS or Binary Ordinal Search distribution). Model inference relies on a Stochastic EM algorithm coupled with a Gibbs sampler, and the ICL-BIC criterion is used for selecting the number of co-clusters (or blocks). The main advantage of this ordinal dedicated co-clustering model is its parsimony, the interpretability of the co-cluster parameters (mode, precision) and the possibility to take into account missing data. Numerical experiments on simulated data show the efficiency of the inference strategy, and real data analyses illustrate the interest of the proposed procedure. The resulting work is now published in the international journal [18]. This is joint work Julien Jacques from University of Lyon 2.

7.3. Axis 1: Model-Based Co-clustering for Ordinal Data of different dimensions

Participant: Christophe Biernacki.

This work has been motivated by a psychological survey on women affected by a breast tumor. Patients replied at different moments of their treatment to questionnaires with answers on ordinal scale. The questions relate to aspects of their life called dimensions. To assist the psychologists in analyzing the results, it is useful to emphasize a structure in the dataset. The clustering method achieves that by creating groups of individuals that are depicted by a representative of the group. From a psychological position, it is also useful to observe how questions may be grouped. This is why a clustering should also be performed on the features, which is called a co-clustering problem. However, gathering questions that are not related to the same dimension does not make sense from a psychologist stance. Therefore, the present work corresponds to perform a constrained co-clustering method aiming to prevent questions from different dimensions from getting assembled in a same column-cluster. In addition, evolution of co-clusters along time has been investigated. The method relies on a constrained Latent Block Model embedding a probability distribution for ordinal data. Parameter estimation relies on a Stochastic EM-algorithm associated to a Gibbs sampler, and the ICL-BIC criterion is used for selecting the numbers of co-clusters. The resulting work is now under revision in an international journal [54] and has been presented to an international conference [38]. The related R package ordinalClust has been also written and has led to a specific preprint [57].

This is joint work with Margot Selosse (PhD student) and Julien Jacques, both from University of Lyon 2, and Florence Cousson-Gélie from University Paul Valéry Montpellier 3.

7.4. Axis 1: Model-based co-clustering for mixed type data

Participant: Christophe Biernacki.

Over decades, a lot of studies have shown the importance of clustering to emphasize groups of observations. More recently, due to the emergence of high-dimensional datasets with a huge number of features, co-clustering techniques have emerged and proposed several methods for simultaneously producing groups of observations and features. By synthesizing the dataset in blocks (the crossing of a row-cluster and a column-cluster), this technique can sometimes summarize better the data and its inherent structure. The Latent Block Model (LBM) is a well-known method for performing a co-clustering. However, recently, contexts with features of different types (here called mixed type datasets) are becoming more common. Unfortunately, the LBM is not directly applicable on this kind of dataset. The present work extends the usual LBM to the so-called Multiple Latent Block Model (MLBM) which is able to handle mixed type datasets. The inference is done through a Stochastic EM-algorithm embedding a Gibbs sampler and model selection criterion is defined to choose the number of row and column clusters. This method was successfully used on simulated and real datasets. This work is available as a preprint [55] which has been submitted to an international journal. It has also led to the R package mixedClust which has been presented to an international workshop [56] and has led to a specific preprint [56].

An adaptation of this general principle to the specific case of mixing textual and continuous data has been also proposed and presented to a national conference [26], with an international audience.

This is joint work with Margot Selosse (PhD student) and Julien Jacques, both from University of Lyon 2.

7.5. Axis 1: Model-Based Co-clustering with Co-variables

Participant: Serge Iovleff.

This work has been motivated by an epidemiological and genetic survey of malaria disease in Senegal. Data were collected between 1990 and 2008. It is based on a latent block model taking into account the problem of grouping variables and clustering individuals by integrating information given by a set of co-variables. Numerical experiments on simulated data sets and an application on real genetic data highlight the interest of this approach. BEM algorithm is deduced and implemented in R package *simerge* and has led to a specific preprint [24].

7.6. Axis 1: Relaxing the Identically Distributed Assumption in Gaussian Co-Clustering for High Dimensional Data

Participant: Christophe Biernacki.

A co-clustering model for continuous data that relaxes the identically distributed assumption within blocks of traditional co-clustering is presented. The proposed model, although allowing more flexibility, still maintains the very high degree of parsimony achieved by traditional co-clustering. A stochastic EM algorithm along with a Gibbs sampler is used for parameter estimation and an ICL criterion is used for model selection. Simulated and real datasets are used for illustr1ation and comparison with traditional co-clustering. This work has led to a preprint

This is a joint work with Michael Gallaugher (PhD student) and Paul Mcnicholas, both from McMaster University (Canada). Michael Gallaughter visited the Modal for three months in 2018.

7.7. Axis 1: Gaussian-based visualization of Gaussian and non-Gaussian model-based clustering

Participants: Christophe Biernacki, Vincent Vandewalle.

A generic method is introduced to visualize in a Gaussian-like way, and onto R^2, results of Gaussian or non-Gaussian model-based clustering. The key point is to explicitly force a spherical Gaussian mixture visualization to inherit from the within cluster overlap which is present in the initial clustering mixture. The result is a particularly user-friendly draw of the clusters, allowing any practitioner to have a thorough overview of the potentially complex clustering result. An entropic measure allows us to inform of the quality of the drawn overlap, in comparison to the true one in the initial space. The proposed method is illustrated on four real data sets of different types (categorical, mixed, functional and network) and is implemented on the R package ClusVis. This work has been submitted to an international journal [12] and has also been presented to an international conference [41].

This is a joint work with Matthieu Marbac from ENSAI.

7.8. Axis 1: A targeted multi-partitions clustering

Participants: Christophe Biernacki, Vincent Vandewalle.

Clustering is generally not a purpose by itself, because its results are mainly tools used by the statistician for another analysis. Indeed, in many applications, clusters are assessed from a set of observed variables, then these clusters are used to predict other variables which are used or not in clustering. Because the final objective of prediction is not considered during cluster analysis, there is no reason to obtain relevant clusters for the variables to predict. We present a unified approach which simultaneously performs cluster analysis and prediction. This method considers that the variables to clusters arise from a product of finite mixture models which provides multiple partition. Moreover, the variables to predict are considered to be independent of the variables to cluster given the partition. The predictions are achieved by a generalized linear model. Model selection is conducted by optimizing the BIC. This optimization is achieved with a modified version of the EM algorithm which performs model selection and maximum likelihood inference simultaneously. An early version of this work has been presented to an international conference [37].

It is a joint work with Matthieu Marbac from ENSAI and with Mohamed Sedki from Université Paris-Sud.

7.9. Axis 1: Co-clustering: A versatile way to perform clustering in high dimension

Participant: Christophe Biernacki.

Standard model-based clustering is known to be very efficient for low-dimensional data sets, but it fails for properly addressing high dimension (HD) ones, where it suffers from both statistical and computational drawbacks. In order to counterbalance this curse of dimensionality, some proposals have been made to take into account redundancy and features utility, but related models are not suitable for too many variables. We advocate that co-clustering, an unsupervised mixture model learning method to define simultaneously groups of rows (individuals) and groups of columns (variables) on a data matrix, is of particular interest to perform HD clustering of individuals even if it is not its primary mission. Indeed, column clustering is recast as a strategy to control the variance of the estimation, the model dimension being driven by the number of groups of variables instead of the number of variables itself. However, the statistical counterpart of this important variance reduction brings naturally some important model bias. The purpose is to access (first in an empirical manner) the trade-off bias-variance of the co-clustering strategy in scenarios involving HD fundaments (correlated variables, irrelevant variables). We show the ability of co-clustering to outperform simple mixture row-clustering, even if co-clustering clearly corresponds to a misspecified model situation, revealing a promising manner to efficiently address (very) HD clustering. An early version of this work has been presented to an international conference [36].

It is a joint work with Christine Keribin from Université Paris-Sud.

7.10. Axis 1: Dealing with missing data in model-based clustering through a MNAR model

Participants: Christophe Biernacki, Fabien Laporte.

Since the 90s, model-based clustering is largely used to classify data. Nowadays, with the increase of available data, missing values are more frequent. Traditional ways to deal with them consist in obtaining a filled data set, either by discarding missing values or by imputing them. In the first case, some information is lost; in the second case, the final clustering purpose is not taken into account through the imputation step. Thus, both solutions risk to blur the clustering estimation result. Alternatively, we defend the need to embed the missingness mechanism directly within the clustering modeling step. There exists three types of missing data: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). In all situations logistic regression is proposed as a natural and flexible candidate model. In particular, its flexibility property allows us to design some meaningful parsimonious variants, as dependency on missing values or dependency on the cluster label. In this unified context, standard model selection criteria can be used to select between such different missing data mechanisms, simultaneously with the number of clusters. Practical interest of our proposal is illustrated on data derived from medical studies suffering from many missing data. An early version of this work has been presented to an international conference [33].

It is a joint work with Gilles Celeux from Inria Saclay and Julie Josse from Ecole Polytechnique.

7.11. Axis 1: Self Organizing Coclustering for textual data synthesis

Participant: Christophe Biernacki.

Recently, different studies have demonstrated the interest of co-clustering, which simultaneously produces clusters of lines and columns. The present work introduces a novel co-clustering model for parsimoniously summarizing textual data in documents × terms format. Besides highlighting homogeneous coclusters - as other existing algorithms do - we also distinguish noisy coclusters from significant ones, which is particularly useful for sparse documents × term matrices. Furthermore, our model proposes a structure among the significant coclusters and thus obtains a better interpretability to the user. By forcing a structure through row-clusters and column-clusters, this approach is competitive in terms of documents clustering, and offers user-friendly results. The algorithm derived for the proposed method is a Stochastic EM algorithm embedding a Gibbs sampling step and the Poisson distribution. A preprint is currently in progress.

This is joint work with Margot Selosse (PhD student) and Julien Jacques, both from University of Lyon 2.

7.12. Axis 1: Linking canonical and spectral clustering

Participants: Christophe Biernacki, Vincent Vandewalle.

It is a recent work aiming at defining a mathematical bridge between classical model-based clustering and classical spectral clustering. Interest of such a prospect is to be able to compare both methods through the rigorous scheme of model selection paradigm. It is an ongoing work.

It is a joint work with Alexandre Lourme from University of Bordeaux.

7.13. Axis 1: Multiple partition clustering

Participant: Vincent Vandewalle.

In the framework of model-based clustering, a model allowing several latent class variables have been proposed. This model assumes that the distribution of the observed data can be factorized into several independent blocks of variables. Each block is assumed to follow a latent class model (i.e., mixture with conditional independence assumption). The proposed model includes variable selection, as a special case, and is able to cope with the mixed-data setting. The simplicity of the model allows to estimate the repartition of the variables into blocks and the mixture parameters simultaneously, thus avoiding running EM algorithms for each possible repartition of variables into blocks. For the proposed method, a model is defined by the number of blocks, the number of clusters inside each block and the repartition of variables into blocks. Model selection can be done with two information criteria, the BIC and the MICL, for which an efficient optimization is proposed. The proposed method gives a rich interpretation of the data set at hand (i.e., analysis of the repartition of the variables into blocks and analysis of the clusters produced by each block of variables). This work as been presented in several international conferences and is now published [20].

It is a joint work with Matthieu Marbac from ENSAI.

7.14. Axis 2: Change-point detection by means of reproducing kernels

Participant: Alain Celisse.

Classical offline change-point detection approaches are limited to detecting changes arising in the mean and/or variance of the distribution along the time. Detecting changes in other moments of the distribution is possible, but at the price of stronger (unrealistic) distributional assumptions which are likely to be violated.

Reproducing kernels are a means to detect changes arising in any moments of the distribution along the time, which are not limited to the mean or the variance. One of the main contributions of this work is to provide a theoretically grounded model selection strategy allowing us to detection multiple changes. From additional extensive simulation experiments, it clearly arises that the so-called KCP approach outperforms numerous state-of-the art change-points detection procedures such as E-divisive, PELT, ...

7.15. Axis 2: New efficient algorithms for multiple change-point detection with kernels

Participants: Alain Celisse, Guillemette Marot.

Several statistical approaches based on reproducing kernels have been proposed to detect abrupt changes arising in the full distribution of the observations and not only in the mean or variance. Some of these approaches enjoy good statistical properties (oracle inequality, ...). Nonetheless, they have a high computational cost both in terms of time and memory. This makes their application difficult even for small and medium sample sizes ($n < 10^4$). This computational issue is addressed by first describing a new efficient and exact algorithm for kernel multiple change-point detection with an improved worst-case complexity that is quadratic in time and linear in space. It allows dealing with medium size signals (up to $n \approx 10^5$). Second, a faster but approximation algorithm is described. It is based on a low-rank approximation to the Gram matrix. It is linear in time and space. This approximation algorithm can be applied to large-scale signals ($n \geq 10^6$). These exact and approximation algorithms have been implemented in R and C for various kernels. The computational and statistical performances of these new algorithms have been assessed through empirical experiments. The runtime of the new algorithms is observed to be faster than that of other considered procedures. Finally, simulations confirmed the higher statistical accuracy of kernel-based approaches to detect changes that are not only in the mean. These simulations also illustrate the flexibility of kernel-based approaches to analyze complex biological profiles made of DNA copy number and allele B frequencies. An R package implementing the approach will be made available on github.

7.16. Axis 2: Multi-Layer Group-Lasso

Participants: Alain Celisse, Guillemette Marot.

Multi-Layer Group-Lasso (MLGL) is a new procedure of variable selection in the context of redundancy between explanatory variables, which holds true with high-dimensional data. A sparsity assumption is made that is, only a few variables are assumed to be relevant for predicting the response variable. In this context, the performance of classical Lasso-based approaches strongly deteriorate as the redundancy strengthens. The proposed approach combines variable aggregation and selection in order to improve interpretability and performance. First, a hierarchical clustering procedure provides at each level a partition of the variables into groups. Then, the set of groups of variables from the different levels of the hierarchy is given as input to group-Lasso, with weights adapted to the structure of the hierarchy. At this step, group-Lasso outputs sets of candidate groups of variables for each value of regularization parameter. The versatility offered by MLGL to choose groups at different levels of the hierarchy a priori induces a high computational complexity. MLGL however exploits the structure of the hierarchy and the weights used in group-lasso to greatly reduce the final time cost. The final choice of the regularization parameter – and therefore the final choice of groups – is made by a multiple hierarchical testing procedures. A paper associated to the R package MLGL has been submitted [45].

7.17. Axis 2: Pseudo-Bayesian Learning with Kernel Fourier Transform as Prior

Participants: Pascal Germain, Gael Letarte.

We revisit the kernel random Fourier features (RFF) method through the lens of the PAC-Bayesian theory. While the primary goal of RFF is to approximate a kernel, we look at the Fourier transform as a prior distribution over trigonometric hypotheses. It naturally suggests learning a posterior on these hypotheses. We derive generalization bounds that are optimized by learning a pseudo-posterior obtained from a closed-form expression, and corresponding learning algorithms This work has been accepted for publication at AISTATS 2019 conference [51].

It is a joint work with Emilie Morvant from Université Jean Monnet de Saint-Etienne.

7.18. Axis 2: Decentralized learning with budgeted network load using Gaussian copulas and classifier ensembles

Participant: Benjamin Guedj.

We examine a network of learners which address the same classification task but must learn from different data sets. The learners can share a limited portion of their data sets so as to preserve the network load. We introduce DELCO (standing for Decentralized Ensemble Learning with COpulas), a new approach in which the shared data and the trained models are sent to a central machine that allows to build an ensemble of classifiers. The proposed method aggregates the base classifiers using a probabilistic model relying on Gaussian copulas. Experiments on logistic regressor ensembles demonstrate competing accuracy and increased robustness as compared to gold standard approaches. A companion python implementation can be downloaded at https://github.com/john-klein/DELCO.

Joint work with John Klein, Olivier Colot, Mahmoud Albardan (all from CRIStAL lab, UMR 9189, Univ. Lille. Preprint submitted: [50].

7.19. Axis 2: Sequential Learning of Principal Curves: Summarizing Data Streams on the Fly

Participants: Benjamin Guedj, Le Li.

When confronted with massive data streams, summarizing data with dimension reduction methods such as PCA raises theoretical and algorithmic pitfalls. Principal curves act as a nonlinear generalization of PCA and the present paper proposes a novel algorithm to automatically and sequentially learn principal curves from data streams. We show that our procedure is supported by regret bounds with optimal sublinear remainder terms. A greedy local search implementation that incorporates both sleeping experts and multi-armed bandit ingredients is presented, along with its regret bound and performance on a toy example and seismic data.

Preprint submitted: [47].

7.20. Axis 2: A Quasi-Bayesian Perspective to Online Clustering

Participants: Benjamin Guedj, Le Li.

When faced with high frequency streams of data, clustering raises theoretical and algorithmic pitfalls. We introduce a new and adaptive online clustering algorithm relying on a quasi-Bayesian approach, with a dynamic (i.e., time-dependent) estimation of the (unknown and changing) number of clusters. We prove that our approach is supported by minimax regret bounds. We also provide an RJMCMC-flavored implementation (called PACBO, see https://cran.r-project.org/web/packages/PACBO/index.html) for which we give a convergence guarantee. Finally, numerical experiments illustrate the potential of our procedure.

Joint work with Sébastien Loustau (LumenAI). Paper published in Electronic Journal of Statistics: https://projecteuclid.org/euclid.ejs/1537430425, [19].

7.21. Axis 2: Pycobra: A Python Toolbox for Ensemble Learning and Visualisation

Participants: Benjamin Guedj, Bhargav Srinivasa Desikan.

We introduce pycobra, a Python library devoted to ensemble learning (regression and classification) and visualisation. Its main assets are the implementation of several ensemble learning algorithms, a flexible and generic interface to compare and blend any existing machine learning algorithm available in Python libraries (as long as a predict method is given), and visualisation tools such as Voronoi tessellations. pycobra is fully scikit-learn compatible and is released under the MIT open-source license. pycobra can be downloaded from the Python Package Index (PyPi) and Machine Learning Open Source Software (MLOSS). The current version (along with Jupyter notebooks, extensive documentation, and continuous integration tests) is available at https://github.com/bhargavvader/pycobra and official documentation website is https://modal.lille.inria.fr/pycobra.

Paper published in Journal of Machine Learning Research: http://jmlr.org/papers/v18/17-228.html, [17]. Software submitted to the `scikit-learn-contrib` repository (under review).

7.22. Axis 2: Simpler PAC-Bayesian bounds for hostile data

Participant: Benjamin Guedj.

PAC-Bayesian learning bounds are of the utmost interest to the learning community. Their role is to connect the generalization ability of an aggregation distribution ρ to its empirical risk and to its Kullback-Leibler divergence with respect to some prior distribution π. Unfortunately, most of the available bounds typically rely on heavy assumptions such as boundedness and independence of the observations. This paper aims at relaxing these constraints and provides PAC-Bayesian learning bounds that hold for dependent, heavy-tailed observations (hereafter referred to as hostile data). In these bounds the Kullack-Leibler divergence is replaced with a general version of Csiszár's f-divergence. We prove a general PAC-Bayesian bound, and show how to use it in various hostile settings.

Joint work with Pierre Alquier (ENSAE ParisTech). Paper published in Machine Learning: [11].

7.23. Axis 2: PAC-Bayesian high dimensional bipartite ranking

Participant: Benjamin Guedj.

This paper is devoted to the bipartite ranking problem, a classical statistical learning task, in a high dimensional setting. We propose a scoring and ranking strategy based on the PAC-Bayesian approach. We consider nonlinear additive scoring functions, and we derive non-asymptotic risk bounds under a sparsity assumption. In particular, oracle inequalities in probability holding under a margin condition assess the performance of our procedure, and prove its minimax optimality. An MCMC-flavored algorithm is proposed to implement our method, along with its behavior on synthetic and real-life datasets.

Joint work with Sylvain Robbiano. Paper published in Journal of Statistical Planning and Inference: [16].

7.24. Axis 2: Multiview Boosting by Controlling the Diversity and the Accuracy of View-specific Voters

Participant: Pascal Germain.

We propose a boosting based multiview learning algorithm which iteratively learns i) weights over view-specific voters capturing view-specific information; and ii) weights over views by optimizing a PAC-Bayes multiview C-Bound that takes into account the accuracy of view-specific classifiers and the diversity between the views. We derive a generalization bound for this strategy following the PAC-Bayes theory which is a suitable tool to deal with models expressed as weighted combination over a set of voters. This work has been submitted to an international journal and is available as a preprint [44].

It is a joint work with Emilie Morvant from Université Jean Monnet de Saint-Etienne and with Massih-Reza Amini of Université de Grenoble, and with Anil Goyal affiliated to both institutions.

7.25. Axis 3: Clustering spatial functional data

Participants: Sophie Dabo, Cristian Preda, Vincent Vandewalle.

We propose two approaches for clustering spatial functional data. The first one is the model-based clustering that uses the concept of density for functional random variables. The second one is the hierarchical clustering based on univariate statistics for functional data such as the functional mode or the functional mean. These two approaches take into account the spatial features of the data: two observations that are spatially close share a common distribution of the associated random variables. The two methodologies are illustrated by an application to air quality data. This work will appear in the “Geostatistical Functional Data Analysis: Theory and Methods”. Wiley, 2018. Editors : Jorge Mateu, Ramon Giraldo [39].

7.26. Axis 3: Categorical functional data analysis

Participants: Cristian Preda, Vincent Vandewalle.

We develop and implement techniques for analysis of categorical functional data. Visualization, clustering and regression methods with categorical functional predictor are proposed.

7.27. Axis 4: Real-time Audio Sources Classification

Participants: Christophe Biernacki, Maxime Baelde.

This work addresses the recurring challenge of real-time monophonic and polyphonic audio source classification. The whole power spectrum is directly involved in the proposed process, avoiding complex and hazardous traditional feature extraction. It is also a natural candidate for polyphonic events thanks to its additive property in such cases. The classification task is performed through a nonparametric kernel-based generative modeling of the power spectrum. Advantage of this model is twofold: it is almost hypothesis free and it allows to straightforwardly obtain the maximum a posteriori classification rule of online signals. Moreover it makes use of the monophonic dataset to build the polyphonic one. Then, to reach the real-time target, the complexity of the method can be tuned by using a standard hierarchical clustering preprocessing of sound models, revealing a particularly efficient computation time and classification accuracy trade-off. The proposed method reveals encouraging results both in monophonic and polyphonic classification tasks on benchmark and owned datasets, even in real-time situations. This method also has several advantages compared to the state-of-the-art methods include a reduced training time, no hyperparameters tuning, the ability to control the computation - accuracy trade-off and no training on already mixed sounds for polyphonic classification. This work is now under revision to an international journal [40].

It is a joint work with Raphaël Greff, from the A-Volute company.

7.28. Axis 4: 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 descriptors equal for both learning and test samples. Indeed, for answering 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 [27], with an international audience, and also to an international conference [28].

It is a joint work with Laetitia Jourdan, from University of Lille.

7.29. Axis 4: Supervised multivariate discretization and levels merging for logistic regression

Participants: Christophe Biernacki, Vincent Vandewalle, Adrien Ehrhardt.

For regulatory and interpretability reasons, the logistic regression is still widely used by financial institutions to learn the refunding probability of a loan given the applicants characteristics from historical data. Although logistic regression handles naturally both quantitative and qualitative data, three ad hoc pre-processing steps are usually performed: firstly, continuous features are discretized by assigning factor levels to predetermined intervals; secondly, qualitative features, if they take numerous values, are grouped; thirdly, interactions (products between two different features) are sparsely introduced. By reinterpreting these discretized (resp. grouped) features as latent variables and by modeling the conditional distribution of each of these latent variables given each original feature with a polytomous logistic link (resp. contingency table), a novel model-based resolution of the discretization problem is introduced. Estimation is performed via a Stochastic Expectation-Maximization (SEM) algorithm and a Gibbs sampler to find the best discretization (resp. grouping) scheme w.r.t. any classical logistic regression loss (AIC, BIC, test set AUC, ...). For detecting interacting features, the same scheme is used by replacing the Gibbs sampler by a Metropolis-Hastings algorithm. The good performances of this approach are illustrated on simulated and real data from Credit Agricole Consumer Finance. This work has been presenting to an international conference in statistics [35] and has been also submitting to an international conference in machine learning. [42].

This is a joint work with Philippe Heinrich from University of Lille.

7.30. Axis 4: MASSICCC Platform for SaaS Software Availability

Participant: Christophe Biernacki.

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, the BlockCluster package has been integrated and also a particular attention to provide meaningful graphical outputs (forMixmod,MixtComp and BlockCluster) directly in the web platform itself has led to some specific developments. In 2018, MASSICCC has been presented to a workshop [29]. Currently, a preprint for an international journal dedicated to software is also in progress.

The MASSICCC platform is available here in the web: https://massiccc.lille.inria.fr.

7.31. Axis 4: ClinMine: Optimizing the Management of Patients in Hospital

Participants: Cristian Preda, Vincent Vandewalle.

A better understanding of "patient pathway" thanks to data analysis can lead to better treatments for patients. The ClinMine project, supported by the French National Research Agency (ANR), aims at proposing, from various case studies, algorithmic and statistical models able to handle this type of pathway data, focusing primarily on hospital data.

Case studies, focusing on the integration of temporal data within analysis has been published [14]. First, the hypothesis that some aspects of the patient pathway can be described, even predicted, from the management process of the hospital medical mail is studied. Therefore a specific functional data analysis is driven, and several types of patients have been detected. The second case study deals with the detection of profiles through a biclustering of the patients. The difficulty to simultaneously deal with heterogeneous data, including temporal data is exposed and a method is proposed.

7.32. 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 R2 with a simple reading of the approximation quality, and provides then a novel variant of dimensionality reduction. This work is still under revision in an international journal [48].

It is a joint work with Hiba Alawieh and Nicolas Wicker, both from University of Lille.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts: SEMENCES DE FRANCE

Sophie Dabo-Niang has a contract with the enterprise SEMENCES DE FRANCE, concerning the realisation of a statistical software.

8.2. Bilateral Contracts: Arcelor-Mittal

Participants: Christophe Biernacki, Vincent Vandewalle.

Arcelor-Mittal is a leader company in steel industry. This contract (which began in 2016 and finished in 2018) aims at optimizing predictive maintenance from mixed data (continuous, categorical, functional) provided by multiple sensors disseminated in steel production lines. Several thousands of sensors are simultaneously involved in this study, most of them providing functional (chronological) values.

It is a joint work with Quentin Grimonprez and Vincent Kubicki (InriaTech engineers).

8.3. Bilateral Contracts: Alstom

Participants: Christophe Biernacki, Benjamin Guedj.

Alstom is a world leader company in integrated transport systems. This contract aims at optimizing predictive maintenance from free text annotations provided by maintenance people. The proposal consists in using co-clustering as a way for grouping both maintenance operations and words describing them.

It is a joint work with Etienne Goffinet (InriaTech engineer).

8.4. Bilateral Contracts: Decathlon

Participant: Christophe Biernacki.

Decathlon is a leading sports retailer.

It is a joint work with Etienne Goffinet (InriaTech engineer). The purpose was to propose a innovative method for sales forecast by using complex data they have (mixed data, chronological series, etc.).

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Bilille partnership

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, CHRU, Inria)

FRABIO, FR3688 (Univ. Lille,CNRS)

CBP / GFS (Univ. Lille, CHRU)

TAG (Univ. Lille, CNRS, INSERM, Institut Pasteur de Lille)

U1167 (Univ. Lille, CHRU, INSERM et Institut Pasteur de Lille)

U1011 (Univ. Lille, INSERM)

UMR8198 (Univ. Lille, CNRS)

LIGAN PM (Univ. Lille, CNRS)

BONSAI (Inria, Univ. Lille, CNRS).

These last teams are thus the main partners of Modal concerning biostatistics for bioinformatics. Guillemette Marot is the co-head 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, UMR 9189 (co-head of bilille)

P. Touzet, UMR 8198 (deputy head of bilille)

C. Bellenguez, U1167

M. Figeac, CBP / GFS

D. Hot, TAG

V. Leclère, Insitut Charles Viollette

M. Lensink, UMR 8576.

O. Sand, IFB-Core.

9.1.2. Bilille collaborations

Participants: Guillemette Marot, Vincent Vandewalle.

Guillemette Marot and Vincent Vandewalle have supervised the data analysis part or support in biostatistics tools testing for the following research projects involving 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):

CIIL, C. Faveeuw, Analysis of cytometry data

CIIL, P. Brodin, Analysis of phenotypic screening data

JPARC, J.M. Taymans, Analysis of translation chips

JPARC, M.C. Chartier-Harlin, RNA-Seq meta-analysis

JPARC, A. Vincent, Microarray analysis

UMR 1167, F. Pinet, Analysis of proteomic data.

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 the leader of the platform, Guillemette Marot is thus involved in these networks.

9.2.2. RHU PreciNASH

Participant: Guillemette Marot.

RHU PreciNASH

Acronym: PreciNASH

Project title: Non-alcoholic steato-hepatitis (NASH) from disease stratification to novel therapeutic approaches

Coordinator: F. Pattou

Duration: 5 years

Partners: FHU Integra and Sanofi

Abstract: PreciNASH, project coordinated by Pr. F. Pattou (UMR 859, EGID), aims at better understanding non alcoholic stratohepatitis (NASH) and improving its diagnosis and care. In this RHU, Guillemette Marot supervises a 2 years post-doc, as her team EA 2694 is a member of the FHU Integra. EA 2694 is involved in the WP1 for the development of a clinical-biological model for the prediction of NASH. Other partners of the FHU are UMR 859, UMR 1011 and UMR 8199, these last three teams being part of the labex EGID (European Genomic Institute for Diabetes). Sanofi is the main industrial partner of the RHU PreciNASH. The whole project will last 5 years (2016-2021).

9.2.3. INS2I-CNRS project PEPS JCJC 2018 “PaRaFF”

Participant: Pascal Germain.

Projet PaRaFF: PAC-Bayesian Random Fourier Features

Coordinator: Emilie Morvant, Hubert Curien Lab, University Jean Monnet, Saint-Etienne

Year: 2018

Abstract: In data science, any method is based on a representation of the data. In this project, we study the learning of representation in the context of automatic learning methods called kernel methods. Our analysis is based on the Random Fourier Features, a method of approximating a kernel function based on a combination random attributes (combination defined by a probability distribution on the attributes). We aim to provide a theoretical understanding of this approach via PAC-Bayesian theory, and to propose a representation learning procedure by exploiting the specificities of this theory.

9.2.4. ANR

9.2.4.1. ANR APRIORI

Participants: Benjamin Guedj, Pascal Germain.

APRIORI 2019–2023, ANR PRC

PAC-Bayesian theory and algorithms for deep learning and representation learning.

Main coordinator of the project: Emilie Morvant, Université Jean Monnet.

Funding: 300k EUR.

2 partners - MODAL (Inria LNE), Hubert Curien Lab. (UMR CNRS 5516).

9.2.4.2. ANR BEAGLE

Participants: Benjamin Guedj, Pascal Germain.

BEAGLE 2019–2023, ANR JCJC

PAC-Bayesian theory and algorithms for agnostic learning

Main coordinator of the project: Benjamin Guedj

Funding: 180k EUR

The consortium also includes Pierre Alquier (ENSAE ParisTech), Peter Grünwald (CWI, The Netherlands), Rémi Bardenet (UMR CRIStAL 9189).

9.2.4.3. ANR SMILE

Participants: Christophe Biernacki, Vincent Vandewalle.

SMILE Project-2018-2022

ANR project (ANR SMILE - Statistical Modeling and Inference for unsupervised Learning at LargE-Scale)

Main coordinator of the project: Faicel Chamroukhi, LMNO, Université de Caen

4 partners - MODAL (Inria LNE), LMNO UMR CNRS 6139 (Caen), LMRS UMR CNRS 6085 (Rouen), LIS UMR CNRS 7020 (Toulon).

9.2.4.4. ANR ClinMine

Participants: Cristian Preda, Vincent Vandewalle.

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.

9.2.4.5. ANR TheraSCUD2022

Participant: Guillemette Marot.

Acronym: TheraSCUD2022

Project title: Targeting the IL-20/IL-22 balance to restore pulmonary, intestinal and metabolic homeostasis after cigarette smoking and unhealthy diet

Coordinator: P. Gosset

Duration: 3 years

Partners: CIIL Institut Pasteur de Lille and UMR 1019 INRA Clermont-Ferrand

Abstract: TheraSCUD2022, project coordinated by P. Gosset (Institut Pasteur de Lille), studies inflammatory disorders associated with cigarette smoking and unhealthy diet (SCUD). Guillemette Marot is involved in this ANR project as head of bilille platform, and will supervise 1 year engineer on integration of omic data. The duration of this project is 3 years (2017-2020).

9.2.5. 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.

9.2.6. Other initiatives

Participants: Serge Iovleff, Cristian Preda, Vincent Vandewalle.

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. Cristian Preda and Vincent Vandewalle are also members of the CloHe project.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

Benjamin Guedj and Vincent Vandewalle are involved on the European H2020 porject PERF-AI

Program: H2020

Project acronym: PERF-AI

Project title: Enhance Aircraft Performance and Optimisation through utilisation of Artificial Intelligence

Duration: November 2018 - November 2020.

Coordinator: Safety Line

Other partners: Safety Line

Abstract: PERF-AI will apply Machine Learning techniques on flight data (parametric and non-parametric approaches) to accurately measure actual aircraft performance throughout its lifecycle.

Within current airline operations, both at flight preparation (on-ground) and at flight management (in-air) levels, the trajectory is first planned, then managed by the Flight Management System (FMS) using a single manufacturer's performance model that is the same for every aircraft of the same type, and also on weather forecast that is computed long before the flight. It induces a lack of accuracy during the planning phase with a flight route pre-established at specific altitudes and speeds to optimize fuel burn, from take-off to landing using aircraft performances that are not those of the real aircraft. Also, the actual flight will usually shift from the original plan because of Air Traffic Control (ATC) constraints, adverse weather, wind changes and tactical re-routing, without possibility for the flight crew, either using the FMS or through connected services to tactically recompute the trajectory in order to continuously optimize the flight path. This is in particular due to the limitations of the performance databases that the current systems are using.

Hence, PERF-AI is focusing on identifying adequate machine learning algorithms, testing their accuracy and capability to perform flight data statistical analysis and developing mathematical models to optimize real flight trajectories with respect to the actual aircraft performance, thus, minimizing fuel consumption throughout the flight.

The consortium consists of Safety-Line (FR) and Inria (FR), having full expertise at Aircraft Performance and Data Science, hence, able to fully propose, test and validate different statistical models that will allow to accurately solve some optimization challenges and implement them in an operational environment.

9.3.2. Collaborations with Major European Organizations

Sophie Dabo-Niang is vice-chair of EMS-CDC (European Mathematical Society-Committe of Developing Countries). She is also a member of the executive committee of CIMPA (International Centre of Pure and Applied Mathematics)

Alain Celisse is a member of a one-year EIT European project called SysBooster with ApSys and Nokia.

9.4. International Initiatives

9.4.1. Inria International Labs

IIL CWI-Inria

Associate Team involved in the International Lab:

9.4.1.1. 6PAC

Title: Making Probably Approximately Correct Learning Active, Sequential, Structure-aware, Efficient, Ideal and Safe

International Partner (Institution - Laboratory - Researcher):

CWI (Netherlands) - Machine Learning Group

PI: Benjamin Guedj

Consortium: Peter Grünwald (co-PI), Wouter Koolen (CWI), Emilie Kaufmann (Inria LNE EPI SequeL).

Start year: 2018 (until 2021)

Webpage: https://bguedj.github.io/6pac/index.html

This project roots in statistical learning theory, which can be viewed as the theoretical foundations of machine learning. The most common framework is a setup in which one is given n training examples, and the goal is to build a predictor that would be efficient on new (similar) data. This efficiency should be supported by PAC (Probably Approximately Correct) guarantees, e.g. upper bounds on the excess risk of a predictor that hold with high probability. Such guarantees however often hold under stringent assumptions which are typically never met in real-life application, e.g., independent, identically distributed data. More realistic modelling of data has triggered many research efforts in several directions: first, accommodating possible data (e.g., dependent, heavy-tailed), and second, in the direction of sequential learning, in which the predictor can be built on the fly, while new data is gathered. We believe that an ever more realistic paradigm is active learning, a setup in which the learner actively requests data (possibly facing constraints, such as storage, velocity, cost, etc.) and adapts its queries to optimize its performance. The 3-years objective of 6PAC (where 6 stands for Sequential, Active, Efficient, Structured, Ideal, Safe - the six research directions we intend to contribute to) is to pave the way to new PAC generalization and sample-complexity upper and lower bounds beyond batch learning. Our ambition is to contribute to several learning setups, ranging from sequential learning (where data streams are collected) to adaptive and active learning (where data streams are requested by the learning algorithm).

9.4.2. Participation in Other International Programs

Starting December 2018, Benjamin Guedj is on sabbatical leave at University College London, Computer Science department, to lead a research team within the UCL AI center.

9.4.2.1. SIMERGE

Title: Statistical Inference for the Management of Extreme Risks and Global Epidemiology

International Partner (Institution - Laboratory - Researcher):

UGB (Senegal) - LERSTAD - Abdou Ka Diongue

Serge Iovleff and Sophie Dabo-Niang are associated members of SIMERGE.

9.5. International Research Visitors

9.5.1. Visits to International Teams

9.5.1.1. Research Stays Abroad

Sophie Dabo-Niang visited the University of Kuala Lumpur, November 2018, the University of Melbourne (Australia), December 2018, and the University of Nador (Morroco), end of December 2018.

Serge Iovleff visited several institutions in Senegal (April 22 – May 18, 2018). He gave a lecture at University Gaston Berger (UGB) of Saint-Louis and collaborated with Cheikh Loucoubar, Seydou Nourou Sylla and Cheikh Loucoubar of the team G4BBM of the Pasteur Institute of Dakar.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Pascal Germain and Benjamin Guedj are the organizers of the https://modal.lille.inria.fr/wikimodal/doku.php?id=seminarsModal team scientific seminar.
Sophie Dabo-Niang is a co-organizer of the 2nd Conference on Econometrics for Environment.
Sophie Dabo-Niang, Cristian Preda and Vincent Vandewalle are the organizers of a session on “Functional Data Analysis” for the conference COMPSTAT 2018.
Vincent Vandewalle is the organizer of a session on advances in model based clustering for the conference ERCIM 2018.

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Sophie Dabo-Niang is the chair of the Scientific Committee of CIMOM18.

10.1.2.2. Member of the Conference Program Committees

Christophe Biernacki is a member of the program committee of MBC2, an international workshop on Model-Based Clustering and Classification (http://mbc2.unict.it).

Cristian Preda was a member of the Scientific Committee of the 9th International Workshop on Applied Probability, IWAP 2018, 18-21 June Budapest, Hungary (https://iwap2018.com).

10.1.2.3. Reviewer

Pascal Germain acted as a reviewer for NIPS 2018, ICML 2018, ICLR 2018, CAp 2018.
Benjamin Guedj served as a reviewer for the top-tier conferences in machine learning ALT 2018, AISTATS 2018, NIPS 2018, ICML 2018, ICLR 2018. He also served as a reviewer for journals (Electronic Journal of Statistics, Journal of Machine Learning Research).
Sophie Dabo-Niang acted as a reviewer for JNP, JSPI, JRSSB, Spatial Statistics, Journal of SFDS, JMVA, ...
Christophe Biernacki acted as a reviewer for a dozen international statistical journals (CSDA, STCO, JMLR, IEEE PAMI, ...).
Serge Iovleff acted as a reviewer for Journal of Statistics and Computing.
Vincent Vandewalle acted as a reviewer for Statistics in Medicine, ADAC and Journal de la SFdS.
Alain Celisse acted as a reviewer for the Annals of Statistics, Bernoulli, JMLR, EJS, JSPI, Artificial Intelligence, ...
Cristian Preda acted as a reviewer for TEST, MCAP, JASA and Bernoulli.

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

Christophe Biernacki is an Associate Editor of the North-Western European Journal of Mathematics (NWEJM) and for Frontiers on the topic “Computational Methods for Data Analytics”. He is also a Guest Editor for the Special Issue on Innovations in Model-Based Clustering and Classification of the journal Advances Data Analysis and Classification (ADAC).

Cristian Preda is an Associate editor of the Methodology and Computing in Applied Probability Journal.

10.1.4. Invited Talks

Christophe Biernacki gave several invited talks:

- One talk at the international conference Compstat 2018
- Four talks at the Summer School on Clustering, Data Analysis and Visualization of Complex Data, May 2018, Catania, Italy [21] [22] [23] [30]
- Two talks at the Research Summer School on Statistics for Data Science – S4D, June 15th-22th 2018, Caen, France [32] [31]
- One talk at the international conference ERCIM 2018 [33]

Vincent Vandewalle:

- Invited talk at the international conference Compstat 2018 [25]
- Invited talk at the international conference ERCIM 2018
- Seminar of the EA 2694 (Université de Lille), Lille, France

Alain Celisse:

- ERCIM, Pise, 15 December 2018
- IWAP, Budapest, July 2018
- WeierstraßInstitute, Berlin, 2018

Pascal Germain:

- Journée Lilloise de Probabilité et Statistiques, Lille, France, 22 June, 2018
- Séminaire de l'équipe PS, Laboratoire Painlevé (University of Lille), Villeneuve d'Ascq France, 9 May, 2018

Benjamin Guedj:

- December 2018, GreekStochastics κ , Athens, Greece
- December 2018, 11th International Conference of the ERCIM working group on Computational and Methodological Statistics (CMStatistics 2018) (invited talk), Pisa, Italy
- September 2018, 2nd Italian-French Statistics Seminar (invited talk), Grenoble, France
- June 2018, 2nd annual congress of the French Mathematical Society (invited talk), Lille, France

Cristian Preda:

- One-Dimensional Discrete Scan Statistics Associated to Some Dependent Models, 5th Stochastic Modeling Techniques and Data Analysis International Conference. 12 - 16 June 2018, Chania, Crete, Greece

10.1.5. Leadership within the Scientific Community

Sophie Dabo-Niang is vice-chair of EMS-CDC.

Till May 2018, Christophe Biernacki was the president of the data mining and learning group of the French statistical association (SFdS, http://www.sfds.asso.fr/).

Since May 2018, Benjamin Guedj has served as president of the Machine Learning and Artificial Intelligence group (MALIA) of the French Statistical association (SFdS, http://www.sfds.asso.fr/).

Since 2017, Benjamin Guedj has been serving as a member of the boards of SFdS and AMIES.

10.1.6. Scientific Expertise

Guillemette Marot reviewed one project as an expert for the ANR and another one for ANSES.

10.1.7. Research Administration

Sophie Dabo-Niang is in charge of the MeQAME axis of the laboratory LEM, CNRS 9221.

Christophe Biernacki has been “Délégué Scientifique” of the Inria Lille center since June 2017.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Sophie Dabo-Niang is teaching

- Master: Spatial Statistics, 24h, M2, University of Lille, France
- Master: Advanced Statistics, 24h, M2, University of Lille, France
- Master: Multivaraite Data Analyses, 24h, M2, University of Lille, France
- Licence: Probability, 24h, L2, University of Lille, France
- Licence: Multivariate Statistics, 24h, L3, University of Lille, France

Guillemette Marot is teaching:

- Licence: Biostatistics, 12h, L1, University of Lille Droit et Santé, France
- Licence: Health care Informatics, 24h, L2, University of Lille Droit et Santé, France
- Master: Biostatistics, 45h, M1, University of Lille Droit et Santé, France
- Master: Supervised classification, 20h, M1, Polytech Lille, France
- Doctorat: Data analysis with R, 7h, University of Lille Droit et Santé, France
- Doctorat: RNA-Seq analysis, 12h, University of Lille Droit et Santé, France

Serge Iovleff is teaching

- Licence: Mathématiques discrètes, 68h, University of Lille, DUT Informatique
- Licence: Modélisation mathématique, 14h, University of Lille, DUT Informatique
- Licence: Algèbre linéaire, 32h, University of Lille, DUT Informatique
- Licence: Analyse et méthodes numériques, 56h, University of Lille, DUT Informatique
- Licence: R.O. et aide à la décision, 32h, University of Lille, DUT Informatique
- Formation Continue: Modélisation, 10h, University of Lille, DUT Informatique
- Master: Introduction to statistics, 16h, University of Lille

Cristian Preda is teaching

- Licence: Linear Regression, 24h, L3, University of Lille, France
- Master: Advanced Statistics, 24h, M1, University of Lille, France
- Master: Biostatistics, 10h, M2, University of Lille, France
- Master: Experimental Designs, 24h, M2, University of Lille, France

Alain Celisse is teaching

- Licence: Graphes et langages, 24h, L3, University of Lille, France
- Licence: Probabilités et statistique, 136h, L3, University of Lille, France
- Formation continue: Probabilités et statistique, 32h, L3, University of Lille, France

Pascal Germain is teaching

- Master: Introduction aux réseaux de neurones, 30h, M2, University of Lille, France

Benjamin Guedj is teaching

- Master: Bayesian Learning, 10h, M2, Centrale Lille, France

10.2.2. Supervision

PhD in progress: Yaroslav Averyanov, November 2017, supervision: Alain Celisse.

PhD in progress: Anne-Lise Bedenel, June 2015, supervision: Christophe Biernacki, Laetitia Jourdan.

PhD in progress: Adrien Ehrhardt, June 2016, supervision: Christophe Biernacki, Philippe Heinrich and Vincent Vandewalle.

PhD defended: Le Li, November 2014–November 2018, supervision: Benjamin Guedj.

PhD in progress: Arthur Leroy, November 2017, supervision: Benjamin Guedj.

PhD in progress: Margot Selosse, October 2017, Christophe Biernacki and Julien Jacques.

PhD in progress: Maxime Baelde, January 2016, Christophe Biernacki and Raphaël Greff.

PhD in progress: Hélène Sarter, Outils statistiques pour la sélection de variables et l'intégration de données “cliniques" et "omiques” : développement et application au registre EPIMAD, December 2016, Corinne Gower and Guillemette Marot.

10.2.3. Juries

Pascal Germain was an examiner at the PhD defense of Valentina Zantedeschi, University Jean Monnet of Saint-Etienne, December 18, 2018.
Benjamin Guedj served as a jury member for the PhD defense of Mahmoud Albardan, Univ. Lille, October 2018.
Sophie Dabo-Niang was a referee at the HDR defense of Tristan Kenga Kiesse, University of Rennes 1, October 22, 2018.
Sophie Dabo-Niang was a referee at the PhD defense of Ousmane Cisse, University Paris 1, December 11th, 2018, and of Julien Ndrin, University of Abidjan (Côte-d'Ivoire), April, 19th, 2018.
Sophie Dabo-Niang was a referee of the PhD dissertation of Javier Álvarez Lièbana, University of Granada (Spain), April, 2018.
Christophe Biernacki acted as a reviewer for PhD theses and one HdR defense. He also acted as an examinator for one PhD thesis and for one HdR defense.

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Guillemette Marot is responsible of bilille, the bioinformatics and bioanalysis platform of Lille. More information about the platform is available at https://wikis.univ-lille.fr/bilille/.

Benjamin Guedj is an appointed deputy member of CLHSCT (Inria LNE).

Benjamin Guedj is an elected member of the Evaluation Committee (CE, Inria).

10.3.2. Internal action

- Pascal Germain gave a talk at the “30 minutes de sciences” seminar of Inria Lille (21/03/2018).
- Pascal Germain gave a talk vulgarizing his work at the “Café des sciences” of Inria Rocquencourt (13/11/2018).
- Guillemette Marot gave a talk at the “30 minutes de sciences” seminar of Inria Lille (21/12/2018).

11. Bibliography

Major publications by the team in recent years

[1] P. ALQUIER, B. GUEDJ. *Simpler PAC-Bayesian Bounds for Hostile Data*, in "Machine Learning", 2018 [*DOI :* 10.1007/s10994-017-5690-0], https://hal.inria.fr/hal-01385064

[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, J. JACQUES.*Model-Based Clustering of Multivariate Ordinal Data Relying on a Stochastic Binary Search Algorithm*, in "Statistics and Computing", 2016, vol. 26, n° 5, p. 929-943, https://hal.inria.fr/hal-01052447

[4] C. BIERNACKI, A. LOURME.*Unifying Data Units and Models in (Co-)Clustering*, in "Advances in Data Analysis and Classification", May 2018, vol. 12, n° 41, https://hal.archives-ouvertes.fr/hal-01653881

[5] A. CELISSE.*Optimal cross-validation in density estimation with the L2-loss*, in "The Annals of Statistics", 2014, vol. 42, n° 5, p. 1879–1910, https://hal.archives-ouvertes.fr/hal-00337058

[6] S. DABO-NIANG, C. TERNYNCK, A.-F. YAO.*Nonparametric prediction in the multivariate spatial context*, in "Journal of Nonparametric Statistics", 2016, vol. 28, n° 2, p. 428-458 [*DOI :* 10.1080/10485252.2016.01.007], https://hal.inria.fr/hal-01425932

[7] J. DUBOIS, V. DUBOIS, H. DEHONDT, P. MAZROOEI, C. MAZUY, A. A. SÉRANDOUR, C. GHEERAERT, P. GUILLAUME, E. BAUGÉ, B. DERUDAS, N. HENNUYER, R. PAUMELLE, G. MAROT, J. S. CARROLL, M. LUPIEN, B. STAELS, P. LEFEBVRE, J. EECKHOUTE.*The logic of transcriptional regulator recruitment architecture at cis -regulatory modules controlling liver functions*, in "Genome Research", June 2017, vol. 27, n° 6, p. 985 - 996 [*DOI :* 10.1101/GR.217075.116], https://hal.archives-ouvertes.fr/hal-01647846

[8] 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

[9] M. MARBAC, V. VANDEWALLE.*A tractable Multi-Partitions Clustering*, in "Computational Statististics & Data Analysis", July 2018 [*DOI :* 10.1016/J.CSDA.2018.06.013], https://hal.inria.fr/hal-01691417

[10] C. PREDA, A. DERMOUNE.*Parametrizations, fixed and random effects*, in "Journal of Multivariate Analysis", February 2017, vol. 154, p. 162 - 176 [*DOI :* 10.1016/J.JMVA.2016.11.001], https://hal.archives-ouvertes.fr/hal-01655461

Publications of the year

Articles in International Peer-Reviewed Journal

[11] P. ALQUIER, B. GUEDJ.*Simpler PAC-Bayesian Bounds for Hostile Data*, in "Machine Learning", 2018 [*DOI :* 10.1007/s10994-017-5690-0], https://hal.inria.fr/hal-01385064

[12] C. BIERNACKI, A. LOURME.*Unifying Data Units and Models in (Co-)Clustering*, in "Advances in Data Analysis and Classification", May 2018, vol. 12, n° 41, https://hal.archives-ouvertes.fr/hal-01653881

[13] S. CURCEAC, C. TERNYNCK, T. B. OUARDA, F. CHEBANA, S. DABO-NIANG.*Short-term air temperature forecasting using Nonparametric Functional Data Analysis and SARMA models*, in "Environmental Modelling and Software", January 2019, vol. 111, p. 394-408 [*DOI :* 10.1016/J.ENVSOFT.2018.09.017], https://hal.inria.fr/hal-01948928

[14] C. Dhaenens, J. Jacques, V. Vandewalle, M. Vandromme, E. Chazard, C. Preda, A. Amarioarei, P. Chaiwuttisak, C. Cozma, G. Ficheur, M.-E. Kessaci, R. Perichon, J. Taillard, R. Bordet, A. Lansiaux, L. Jourdan, D. Delerue, A. Hansske.*ClinMine: Optimizing the Management of Patients in Hospital*, in "IRBM", January 2018, vol. 39, n^{o} 2, p. 83-92 [*DOI :* 10.1016/j.irbm.2017.12.002], https://hal.inria.fr/hal-01692197

[15] R. Giraldo, S. Dabo-Niang, S. Martinez.*Statistical modeling of spatial big data: An approach from a functional data analysis perspective*, in "Statistics & Probability Letters", February 2018, vol. 136, p. 126-129 [*DOI :* 10.1016/j.spl.2018.02.025], https://hal.archives-ouvertes.fr/hal-01744181

[16] B. Guedj, S. Robbiano.*PAC-Bayesian High Dimensional Bipartite Ranking*, in "Journal of Statistical Planning and Inference", 2018 [*DOI :* 10.1016/j.jspi.2017.10.010], https://hal.inria.fr/hal-01226472

[17] B. Guedj, B. Srinivasa Desikan.*Pycobra: A Python Toolbox for Ensemble Learning and Visualisation*, in "Journal of Machine Learning Research", June 2018, vol. 18, p. 1 - 5, https://hal.inria.fr/hal-01514059

[18] J. Jacques, C. Biernacki.*Model-Based Co-clustering for Ordinal Data*, in "Computational Statistics & Data Analysis", July 2018, vol. 123, 15, https://hal.inria.fr/hal-01448299

[19] L. Li, B. Guedj, S. Loustau.*A Quasi-Bayesian Perspective to Online Clustering*, in "Electronic journal of statistics ", 2018 [*DOI :* 10.1214/18-EJS1479], https://hal.inria.fr/hal-01264233

[20] M. Marbac, V. Vandewalle.*A tractable Multi-Partitions Clustering*, in "Computational Statistics & Data Analysis", July 2018 [*DOI :* 10.1016/j.csda.2018.06.013], https://hal.inria.fr/hal-01691417

Invited Conferences

[21] C. Biernacki.*Going further in cluster analysis and classification: Bi-clustering and co-clustering*, in "Summer School on Clustering, Data Analysis and Visualization of Complex Data", Catania, Italy, May 2018, https://hal.inria.fr/hal-01810380

[22] C. Biernacki.*Introduction to cluster analysis and classification: Evaluating clustering*, in "Summer School on Clustering, Data Analysis and Visualization of Complex Data", Catania, Italy, May 2018, https://hal.inria.fr/hal-01810377

[23] C. Biernacki.*Introduction to cluster analysis and classification: Performing clustering*, in "Summer School on Clustering, Data Analysis and Visualization of Complex Data", Catania, Italy, May 2018, https://hal.inria.fr/hal-01810376

[24] S. Iovleff, S. N. Sylla.*blockcluster, simerge and C++ with R*, in "Mixture Models: Theory and Applications", Paris, France, June 2018, https://hal.inria.fr/hal-01884822

[25] V. Vandewalle, M. Marbac.*A tractable multi-partitions clustering*, in "COMPSTAT 2018 - 23rd International Conference on Computational Statistics", Iasi, Romania, August 2018, https://hal.inria.fr/hal-01956922

National Conferences with Proceeding

[26] M. Selosse, J. Jacques, C. Biernacki.*Co-clustering de données textuelles et continues*, in "SFdS 2018 - 50èmes Journées de Statistique", Saclay, France, May 2018, https://hal.inria.fr/hal-01797493

Conferences without Proceedings

[27] A.-L. BEDENEL, L. JOURDAN, C. BIERNACKI.*Probabilities estimation by a genetic algorithm*, in "ROADEF2018", Lorient, France, February 2018, https://hal.archives-ouvertes.fr/hal-01868195

[28] A.-L. BEDENEL, L. JOURDAN, C. BIERNACKI.*Probability estimation by an adapted genetic algorithm in web insurance*, in "LION 12 - Learning and Intelligent Optimization Conference", Kalamata, Greece, June 2018, https://hal.archives-ouvertes.fr/hal-01885117

[29] C. BIERNACKI, B. AUDER, G. CELEUX, J. DEMONT, F. LANGROGNET, V. KUBICKI, C. POLI, J. RENAULT.*MASSICCC: A SaaS Platform for Clustering and Co-Clustering of Mixed Data*, in "Workshop MixStatSeq: “Mixture models: Theory and applications”", Paris, France, June 2018, https://hal.archives-ouvertes.fr/hal-01949175

[30] C. BIERNACKI.*Introduction to cluster analysis and classification: Formalizing clustering*, in "Summer School on Clustering, Data Analysis and Visualization of Complex Data", Catania, Italy, May 2018, https://hal.inria.fr/hal-01810379

[31] C. BIERNACKI.*Model selection theory and considerations in large scale scenarios*, in "Research Summer School on Statistics for Data Science – S4D", Caen, France, June 2018, https://hal.archives-ouvertes.fr/hal-01949168

[32] C. BIERNACKI.*Model-based clustering and co-clusteringin high-dimensional scenarios*, in "Research Summer School on Statistics for Data Science – S4D", Caen, France, June 2018, https://hal.archives-ouvertes.fr/hal-01949167

[33] C. BIERNACKI, G. CELEUX, J. JOSSE, F. LAPORTE.*Dealing with missing data in model-based clustering through a MNAR model*, in "CMStatistics 2018 - 11th International Conference of the ERCIM WG on Computational and Methodological Statistics", Pise, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01949120

[34] C. BIERNACKI, V. VANDEWALLE, M. MARBAC.*Gaussian-based visualization of Gaussian and non-Gaussian model-based clustering*, in "23rd International Conference on Computational Statistics", Iasi, Romania, August 2018, https://hal.archives-ouvertes.fr/hal-01949127

[35] A. EHRHARDT, V. VANDEWALLE, C. BIERNACKI, P. HEINRICH.*Supervised multivariate discretization and levels merging for logistic regression*, in "23rd International Conference on Computational Statistics", Iasi, Romania, August 2018, https://hal.archives-ouvertes.fr/hal-01949128

[36] C. KERIBIN, C. BIERNACKI.*Co-clustering: A versatile way to perform clustering in high dimension*, in "The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2018)", Pise, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01949116

[37] M. MARBAC, C. BIERNACKI, M. SEDKI, V. VANDEWALLE.*A targeted multi-partitions clustering*, in "The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2018)", Pise, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01949111

[38] M. SELOSSE, J. JACQUES, C. BIERNACKI.*Analyzing large matrices of ordinal data*, in "The 11th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2018)", Pise, Italy, December 2018, https://hal.archives-ouvertes.fr/hal-01949095

Scientific Books (or Scientific Book chapters)

[39] S. DABO-NIANG, C. PREDA, V. VANDEWALLE.*Clustering spatial functional data*, in "Geostatistical Functional Data Analysis : Theory and Methods. Editors: Jorge Mateu, Ramon Giraldo", 2018, https://hal.inria.fr/hal-01948934

Other Publications

[40] M. BAELDE, C. BIERNACKI, R. GREFF.*Real-Time Monophonic and Polyphonic Audio Classification from Power Spectra*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01834221

[41] C. BIERNACKI, M. MARBAC, V. VANDEWALLE.*Gaussian Based Visualization of Gaussian and Non-Gaussian Based Clustering*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01949155

[42] A. EHRHARDT, C. BIERNACKI, V. VANDEWALLE, P. HEINRICH.*Feature quantization for parsimonious and interpretable predictive models*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01949135

[43] P. GERMAIN, A. HABRARD, F. LAVIOLETTE, E. MORVANT.*PAC-Bayes and Domain Adaptation*, November 2018, https://arxiv.org/abs/1707.05712 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01563152

[44] A. GOYAL, E. MORVANT, P. GERMAIN, M.-R. AMINI.*Multiview Boosting by Controlling the Diversity and the Accuracy of View-specific Voters*, August 2018, https://arxiv.org/abs/1808.05784 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01857463

[45] Q. GRIMONPREZ, S. BLANCK, A. CELISSE, G. MAROT.*MLGL: An R package implementing correlated variable selection by hierarchical clustering and group-Lasso*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01857242

[46] B. GUEDJ.*A Primer on PAC-Bayesian Learning*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01983732

[47] B. GUEDJ, L. LI.*Sequential Learning of Principal Curves: Summarizing Data Streams on the Fly*, May 2018, working paper or preprint, https://hal.inria.fr/hal-01796011

[48] A. HIBA, N. WICKER, C. BIERNACKI.*Projection under pairwise distance controls*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01420662

[49] S. IOVLEFF, S. N. SYLLA, C. LOUCOUBAR.*Block clustering of Binary Data with Gaussian Co-variables*, December 2018, https://arxiv.org/abs/1812.08520 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01961978

[50] J. KLEIN, M. ALBARDAN, B. GUEDJ, O. COLOT.*Decentralized learning with budgeted network load using Gaussian copulas and classifier ensembles*, April 2018, https://arxiv.org/abs/1804.10028 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01779989

[51] G. LETARTE, E. MORVANT, P. GERMAIN.*Pseudo-Bayesian Learning with Kernel Fourier Transform as Prior*, 2018, https://arxiv.org/abs/1810.12683 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01908555

[52] G. MAZO, Y. AVERYANOV.*Constraining kernel estimators in semiparametric copula mixture models*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01774629

[53] A. PODDAR, S. IOVLEFF, F. LATIMIER.*Estimation of Parsimonious Covariance Models for Gaussian Matrix Valued Random Variables for Multi-Dimensional Spectroscopic Data*, December 2018, WiML 2018 - 13th Women in Machine Learning workshop, Poster, https://hal.archives-ouvertes.fr/hal-01954769

[54] M. SELOSSE, J. JACQUES, C. BIERNACKI, F. COUSSON-GÉLIE.*Analyzing quality of life survey using constrained co-clustering model for ordinal data and some dynamic implication*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01643910

[55] M. SELOSSE, J. JACQUES, C. BIERNACKI.*Model-based co-clustering for mixed type data*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01893457

[56] M. SELOSSE, J. JACQUES, C. BIERNACKI.*mixedClust: an R package for mixed data classiication, clustering and co-clustering*, July 2018, 25th Summer Session Working Group on Model-Based Clustering, Poster, https://hal.archives-ouvertes.fr/hal-01949171

[57] M. SELOSSE, J. JACQUES, C. BIERNACKI.*ordinalClust: an R package for analyzing ordinal data*, September 2018, working paper or preprint, https://hal.inria.fr/hal-01678800

Team NON-A POST

Non-Asymptotic estimation for online 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
Optimization and control of dynamic systems

Table of contents

1. Team, Visitors, External Collaborators **291**
2. Overall Objectives **292**
3. Research Program **293**
3.1. Theory of homogeneous systems 293
3.2. Numerical differentiation 294
3.3. FT and FxT control and estimation 294
3.4. Applications 295
4. Highlights of the Year **295**
5. New Software and Platforms **295**
6. New Results **296**
6.1. Implementation of finite- and fixed-time algorithms 296
6.2. A solution to finite- and fixed-time estimation 296
6.3. Numeric and analytic design of homogeneous Lyapunov functions 296
6.4. Distributed finite-time estimation 296
7. Bilateral Contracts and Grants with Industry **296**
8. Partnerships and Cooperations **296**
8.1. Regional Initiatives 296
8.2. National Initiatives 297
8.3. European Initiatives 297
8.4. International Initiatives 297
8.4.1. Inria Associate Teams Not Involved in an Inria International Labs 297
8.4.2. Inria North European Lab 297
8.4.3. Informal International Partners 298
9. Dissemination **298**
9.1.1. Scientific Events Selection 298
9.1.1.1. Member of the Conference Program Committees 298
9.1.1.2. Reviewer 298
9.1.2. Journal 298
9.1.2.1. Member of the Editorial Boards 298
9.1.2.2. Reviewer - Reviewing Activities 298
9.1.3. Research Administration 299
10. Bibliography **299**

Team NON-A POST

Creation of the Team: 2018 January 01, end of the Team: 2018 December 31

Keywords:

Computer Science and Digital Science:

A5.1.1. - Engineering of interactive systems
A5.1.5. - Body-based interfaces
A5.9.1. - Sampling, acquisition
A5.9.2. - Estimation, modeling
A5.10.3. - Planning
A5.10.4. - Robot control
A5.10.6. - Swarm robotics
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization

Other Research Topics and Application Domains:

B3.4.3. - Pollution
B4.5. - Energy consumption
B5.6. - Robotic systems
B6.4. - Internet of things
B6.6. - Embedded systems
B7.1. - Traffic management
B7.1.2. - Road traffic
B7.2.1. - Smart vehicles

1. Team, Visitors, External Collaborators

Research Scientists

Denis Efimov [Team leader, Inria, Researcher, HDR]
Andrey Polyakov [Inria, Researcher, HDR]

Faculty Member

Jean-Pierre Richard [Permanent head, Centrale Lille, Professor, HDR]

External Collaborators

Gerald Dherbomez [CRIStAL, CNRS]
Leonid Fridman [Professor, UNAM, Inria International Chair]

PhD Students

Youness Braidiz [ANR]
Nelson de Figueiredo Barroso [ANR, Region]
Tatiana Kharkovskaia [EC Lille–ITMO]
Edouard Leurent [Inria-Renault]
Francisco Lopez Ramirez [Inria]
Gabriele Perozzi [Onera-Region]
Haik Jan Silm [EC Lille-KUL]

Jijju Thomas [EC Lille-TU Eindhoven]
Quentin Voortman [TU Eindhoven-EC Lille]
Yue Wang [EC Lille]
Siyuan Wang [EC Lille]

Post-Doctoral Fellows
Nicolas Espitia [Inria]
Tonametl Sanchez Ramirez [ANR]

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.
- 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 Internet of things, 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 POST* addresses 3 particular theoretical challenges:

1) Development of theory of dynamical homogeneous systems;

2) Estimate on-line the derivatives of a signal;

3) Design of control and estimation algorithms converging in finite and in fixed time.

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 POST* are within the continuity of the project-teams *Non-A* and *ALIEN* in what concerns the *algebraic tools* that are developed for finite-time estimation purposes. However, *Non-A POST* 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 POST also wants to confront these theoretical challenges with some application fields: Networked robots, Nano/macro machining, quadrotors, active flow control, *etc*. Today, most of our effort (*i.e.*, engineering staff) is devoted to the first item, according to the theme "Internet of Things" (IoT) promoted by Inria in its Strategic Plan for the Lille North-Europe research center.

3. Research Program

3.1. Theory of homogeneous systems

Homogeneity is a property of mathematical objects, such as functions or vector fields, to be scaled in a consistent manner with respect to a scaling operation (called a dilation) applied to their argument (a kind of symmetry). The first rise of homogeneity deals with homogeneous polynomials investigated by L. Euler in 18^{th} century. In 50s and 60s more generic notions of homogeneity (weighted and coordinate-free or geometric) have been introduced by V.I. Zubov and his group. For example, a function $f : \mathbb{R}^n \to \mathbb{R}$ is called homogeneous (in Euler's sense) if

$$f(\lambda x) = \lambda^{1+\nu} f(x) \quad \forall x \in \mathbb{R}^n,\ \forall \lambda > 0$$

for some $\nu \geq -1$ called the degree of homogeneity of f (a parameter of symmetry). Such a type of symmetry leads to the scaling of trajectories of resultant dynamical systems, *e.g.* for

$$\dot{x}(t) = f\left(x(t)\right) \quad t \geq 0,\ x(0) = x_0 \in \mathbb{R}^n$$

denote a solution corresponding to the initial condition x_0 by $X(t, x_0)$, then

$$X(t, \lambda x_0) = \lambda X(\lambda^{\nu} t, x_0) \quad \forall x_0 \in \mathbb{R}^n,\ \forall \lambda > 0.$$

So homogeneous systems possess several important and useful properties: their local behavior is the same as global one, the rate of convergence to the origin can be identified by degree of homogeneity, the stability is robust to various perturbations. There are also plenty of researches performed in the last 30 years and the members of *Non-A POST* team extended these notions of homogeneity to discontinuous systems (in a geometric framework), time-delay systems, partial differential equations, time-varying systems, and recently to discrete-time models (together with the concept of local homogeneity). They also proposed plenty of control and estimation algorithms based on homogeneity.

Advantages of homogeneous algorithms taking into account the above mentioned criteria:

A1) The rate of convergence in homogeneous systems can be qualified by its degree (finite-time and fixed-time for negative and positive degrees, respectively).

A2) Due to symmetry of these systems they admit special discretization tools (also developed by the members of *Non-A POST* team), which make simpler they realization for on-line scenarios.

A3) The internal symmetry of these dynamics makes them inherently robust with respect to external perturbations, measurement noises and delays, which is especially important in networked systems.

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 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* and *Non-A* teams, which rely on the only information contained in the output signal and its derivatives.

3.3. FT and FxT control and estimation

To design an estimation or control algorithm we have to select a performance criterion to be optimized. Stability is one of the main performance indexes, which has to be established during analysis or design of a dynamical system. Stability is usually investigated with respect to an invariant mode (*e.g.*, an equilibrium, desired trajectory or a limit cycle), then another important characteristics is the time of convergence of the system trajectories to this mode, which can be *asymptotic* (in conventional approaches) or *finite-time* (being the focus of *Non-A POST* team). In the latter case the limit mode has to be exactly established in a finite time dependent on initial deviations (if such a time is independent on initial conditions, then this type of convergence is called *fixed-time*). If the rate of convergence is just faster than any exponential of time, then such a convergence is called *hyperexponential*. The notion of finite-time stability has been proposed in 60s by E. Roxin and it has been developed in many works later, where a particular attention is paid to the time of convergence for trajectories to a steady state (it is worth to note that there exists another notion having the same name, *i.e.* finite-time or short-time stability, which is focused on analysis of a dynamical system behavior on bounded intervals of time, and it is completely different and not considered here). For example, the following simple scalar dynamics:

$$\dot{x}(t) = -|x(t)|^{1+\nu}\mathrm{sign}(x(t)) \quad \forall t \geq 0,\ x(t) \in \mathbb{R},$$

has the solution $x(t) = \beta(|x_0|, t)\mathrm{sign}(x_0)$ for any initial condition $x(0) = x_0 \in \mathbb{R}$, where

$$\beta(s,t) = \begin{cases} \begin{cases} \left(s^{-\nu} + \nu t\right)^{-\frac{1}{\nu}} & t < -\frac{s^{-\nu}}{\nu} \\ 0 & t \geq -\frac{s^{-\nu}}{\nu} \end{cases} & -1 \leq \nu < 0 \\ \mathrm{e}^{-t} s & \nu = 0 \\ \frac{s}{(1+\nu s^{\nu} t)^{\frac{1}{\nu}}} & \nu > 0 \end{cases},$$

[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").

[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.

which possesses a finite-time convergence with the settling time $-\frac{|x_0|^{-\nu}}{\nu}$ for $-1 \leq \nu < 0$, an exponential convergence for $\nu = 0$ and the fixed time of convergence to the unit ball is bounded by ν^{-1} for $\nu > 0$. It is straightforward to check that this simple system is homogeneous of degree ν. The members of *Non-A POST* team obtained many results on analysis and design of control and estimation algorithms in this context. A useful and simple method to deal with these three types of convergence (finite-time, fixed-time or hyperexponential) is based on the theory of *homogeneous* systems.

3.4. Applications

The application of the developed control and estimation algorithms for different scenarios in IoT is a priority for *Non-A POST* team. Participation in different potential applications allows the team to better understand the features of IoT and their required performances. A list of possible applications, partially already addressed in the team, is as follows:

- smart bivalve-based biosensor for water quality monitoring (ANR project WaQMoS): presence of persistent external perturbations, which are hard to measure, and important model uncertainty make application of conventional techniques complicated; another issue is consensus seeking between animals for a contamination detection;
- control and estimation for flying vehicles, *e.g.* quadrotors or blimps (1 PhD ONERA, 2 PhDs EC Lille): nonlinearity of the model and its uncertainty coupled with important aerodynamic perturbations have to be compensated by fast (finite- or fixed-time) and robust control and estimation algorithms;
- human behavior modeling and estimation with posterior design of algorithms for human-computer interaction (ANR project TurboTouch): robust finite-time differentiators demonstrate good estimation capabilities needed for prediction in this application;
- human physiological characteristics estimation (like emotion detection, galvanic skin response filtering, fatigue evaluation in collaborations with Neotrope and Ellcie Healthy): intelligent robust filtering and finite-time distributed estimation are key features in this scenario;
- path planning for autonomous vehicles taking into account behavior of humans (PhD CIFRE with SEQUEL team and Renault): application of interval estimation and prediction techniques to treat the uncertainty of the environment by reducing computational complexity of reinforcement learning;
- flow control (in the framework of ContrATech subprogram of CPER ELSAT): the case of control and estimation of a distributed-parameter system with very fast and uncertain dynamics, where finite-time solutions developed by *Non-A POST* team are necessary

Involvement in various real-world scenarios will allow *Non-A POST* to develop demonstrators of disposed technologies with application to IoT.

4. Highlights of the Year

4.1. Highlights of the Year

4.1.1. Awards

- Gabriele Perozzi (a PhD student of the team) get the creativity prize of FR CNRS TTM (La Fédération de Recherche Transports Terrestres & Mobilité);
- Hafiz Ahmed (a former PhD student of the team) is a winner of Annual European PhD Award on Control for Complex and Heterogeneous Systems.

5. New Software and Platforms

5.1. ADHOMFI

Adaptive Homogeneous Filtering

KEYWORDS: Automatic differentiation - Filtering
FUNCTIONAL DESCRIPTION: allows to reconstruct a signal based on derivatives estimation and to filter high amplitude and wide frequencies spectrum perturbations.

- Contact: Denis Efimov

6. New Results

6.1. Implementation of finite- and fixed-time algorithms

In [22] several algorithms of implicit discretization for generalized homogeneous systems having discontinuity only at the origin are developed. They are based on the transformation of the original system to an equivalent one which admits an implicit or a semi-implicit discretization schemes preserving the stability properties of the continuous-time system. Namely, the discretized model remains finite-time stable (in the case of negative homogeneity degree), and practically fixed-time stable (in the case of positive homogeneity degree).

6.2. A solution to finite- and fixed-time estimation

The work [18] deals with the problem of finite-time and fixed-time observation of linear multiple input multiple output (MIMO) control systems. The proposed nonlinear dynamic observers guarantee convergence of the observer states to the original system state in a finite and in a fixed (defined *a priori*) time. Algorithms for the observers parameters tuning are also provided and a robustness analysis against input disturbances and measurement noises is carried out.

6.3. Numeric and analytic design of homogeneous Lyapunov functions

The problem of the synthesis of a homogeneous Lyapunov function for an asymptotically stable homogeneous system is studied in [10]. First, for systems with nonnegative degree of homogeneity, several expressions of homogeneous Lyapunov functions are derived, which depend explicitly on the supremum or the integral (over finite or infinite intervals of time) of the system solutions. Second, a numeric procedure is proposed, which ensures the construction of a homogeneous Lyapunov function.

6.4. Distributed finite-time estimation

In [29] the robust distributed estimation for a class of time-invariant plants is achieved via a finite-time observer, its error reaching zero after a finite time in the absence of perturbation. Two types of robustness are also shown. First, input-to-state stability with respect to measurement noises and additive perturbations is proven. Second, we demonstrate that the estimation error stays bounded in the presence of known transmission delays.

7. Bilateral Contracts and Grants with Industry

7.1. Bilateral Contracts with Industry

A transfer contract with Ellcie Healthy on intelligent filtering of measurements in smart eyeglasses.

8. Partnerships and Cooperations

8.1. Regional Initiatives

Non-A POST team hosts CPER Data ControlHub (an on-line laboratory for control system experimentation) and participates at ContrATech subprogram of CPER ELSAT.

8.2. National Initiatives

- Inria Project Lab (IPL) IPL COSY.
- 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, 2015-2020.
- ANR project TurboTouch (High-performance touch interactions), coordinator: G. Casiez (MJOLNIR team, Inria), 2014-2019.
- ANR project DIGITSLID (DIGITal set-valued and homogeneous SLIding mode control and Differentiators: the implicit approach), coordinator: Bernard Brogliato (TriPOP team, Inria), 2018-2022.
- ANR project ROCC-SYS (Robust Control of Cyber-Physical Systems), coordinator: L. Hetel (CNRS, EC de Lille), 2013-2018.
- 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".

8.3. European Initiatives

8.3.1. FP7 & H2020 Projects

UCoCoS: the objectives of the project are to create a control-oriented framework for complex systems, and to define a common language, common methods, tools and software for the complexity scientist. The principal investigators are: W. Michiels, J.-P. Richard and H. Nijmeijer.

8.4. International Initiatives

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

8.4.1.1. HoTSMoCE

Title: Homogeneity Tools for Sliding Mode Control and Estimation

International Partner (Institution - Laboratory - Researcher):

UNAM (Mexico)

Prof. Leonid Fridman

2016–2018

The team *Non-A POST* is developing an estimation theory, built around differential algebra and operational calculation on the one hand, and high gain algorithms (such as sliding mode) on the other hand. The Mexican partner team comes from "Sliding Mode Control" laboratory of UNAM. There exists a strong intersection of interests of both teams (application of homogeneity for design of sliding mode control and estimation algorithms, and analysis of finite-time stability). That is why there exists a long history of collaboration between these two teams. The goal of the project is development of control and estimation algorithms converging in fixed or in finite time by applying the last generation sliding mode techniques and the homogeneity theory. The project realization is planned in the form of short-time visits of permanent staff and visits of PhD students for a long period of stay. Such visits are very important for young scientists, and also help Non-A team to prepare and find good PhDs/post-docs for future.

8.4.2. Inria North European Lab

RECoT, "Robust Estimation and Control with Time Constraints", 2018–2020

International Partner: IBM Research, Dublin (Dr. Sergiy Zhuk)

Non-A Post team of Inria deals with control and estimation of on-line (dynamical) systems with applications to robotics, biological systems, human-machine interfaces and active ow control. The key feature of the developed algorithms is a robustness and a non-asymptotic convergence allowing to fulfill some time constraints. The main methodology is a homogeneity (dilation symmetry) approach. IBM Research team develops minimax algorithms for state estimation and identification of dynamical systems with applications to computational fluid dynamics and image assimilation problems. The key feature of the resulting algorithms is the exact or approximate description of the reachability set of the underlying dynamical system in finite or infinite dimensions. The methodology is relies upon duality and Lyapunov exponents. The objectives of the collaboration are an exchange of the scientific knowledge and the joint research of the following problems: homogeneous observers design using minimax approach; development of fast and consistent computational algorithms for digital implementation of homogeneous controllers and observers; extension of sliding mode control methodology to infinite-dimensional models using minimax approach; the minimax observer-based control design for turbulent flows.

8.4.3. Informal International Partners

- ITMO University, Saint-Petersburg, Russia
- Tel-Aviv University, Tel-Aviv, Israel
- CINVESTAV-IPN, Mexico, Mexico
- Hangzhou Dianzi University, Hangzhou, China
- Brandenburg University of Technology, Cottbus, Germany

9. Dissemination

9.1. Promoting Scientific Activities

9.1.1. Scientific Events Selection

9.1.1.1. Member of the Conference Program Committees

- Richard J.-P., EUCA-IEEE ECC, Limassol, Cyprus
- Richard J.-P., IFAC TDS, Budapest, Hungary
- Richard J.-P., IARA VEHICULAR, Venice, Italy
- Efimov D., IFAC CHAOS, Eindhoven, Netherlands
- Efimov D., IFAC MICNON, Guadalajara, Mexico

9.1.1.2. Reviewer

The members of the team serve as reviewers to all major conferences in the field: IEEE CDC, ECC, ACC *etc*.

9.1.2. Journal

9.1.2.1. Member of the Editorial Boards

- Polyakov A., International Journal of Robust and Nonlinear Control
- Polyakov A., Journal of Optimization Theory and Applications
- Efimov D., IFAC Journal on Nonlinear Analysis: Hybrid Systems
- Efimov D., Asian Journal of Control
- Efimov D., IEEE Transactions on Automatic Control

9.1.2.2. Reviewer - Reviewing Activities

The members of the team serve as reviewers to all major journals in the field: IEEE Trans. Automatic Control, Automatica, Systems & Control Letters, SIAM Journal on Optimization and Control, Int. Journal of Robust and Nonlinear Control *etc*.

9.1.3. Research Administration

- Richard J.-P., Investigator for the CNRS FR TTM
- Efimov D., Chair of EECI PhD Award

10. Bibliography

Publications of the year

Articles in International Peer-Reviewed Journal

[1] H. AHMED, R. USHIROBIRA, D. EFIMOV.*Robust global synchronization of Brockett oscillators*, in "IEEE Transactions on Control of Network Systems", 2018, p. 1-10 [*DOI :* 10.1109/TCNS.2018.2813927], https://hal.inria.fr/hal-01965539

[2] A. APARICIO, L. FRIDMAN, D. EFIMOV.*Stabilization of systems with switchings on the axis of their coordinates and its input-to-state properties*, in "Nonlinear Analysis: Hybrid Systems", 2018, https://hal.inria.fr/hal-01908289

[3] O. BARA, M. FLIESS, C. JOIN, J. DAY, S. M. DJOUADI.*Toward a model-free feedback control synthesis for treating acute inflammation*, in "Journal of Theoretical Biology", July 2018, vol. 448, p. 26-37 [*DOI :* 10.1016/J.JTBI.2018.04.003], https://hal-polytechnique.archives-ouvertes.fr/hal-01756743

[4] K. H. DEGUE, D. EFIMOV, J.-P. RICHARD.*Stabilization of linear impulsive systems under dwell-time constraints: Interval observer-based framework*, in "European Journal of Control", January 2018 [*DOI :* 10.1016/J.EJCON.2018.01.001], https://hal.inria.fr/hal-01761831

[5] D. DILEEP, R. VAN PARYS, G. PIPELEERS, L. HETEL, J.-P. RICHARD, W. MICHIELS.*Design of robust decentralised controllers for MIMO plants with delays through network structure exploitation*, in "International Journal of Control", November 2018, p. 1-15 [*DOI :* 10.1080/00207179.2018.1554268], https://hal.archives-ouvertes.fr/hal-01951755

[6] B.-H. DU, A. POLYAKOV, G. ZHENG, Q. QUAN.*Quadrotor trajectory tracking by using fixed-time differentiator*, in "International Journal of Control", April 2018 [*DOI :* 10.1080/00207179.2018.1462534], https://hal.inria.fr/hal-01947365

[7] D. EFIMOV, A. POLYAKOV, A. LEVANT, W. PERRUQUETTI.*Convergence acceleration for observers by gain commutation*, in "International Journal of Control", February 2018, p. 1-20, https://hal.inria.fr/hal-01651747

[8] D. EFIMOV, A. POLYAKOV, A. LEVANT, W. PERRUQUETTI.*Realization and Discretization of Asymptotically Stable Homogeneous Systems*, in "IEEE Transactions on Automatic Control", May 2018, https://hal.inria.fr/hal-01514350

[9] D. EFIMOV, J. SCHIFFER.*On boundedness of solutions of state periodic systems: a multivariable cell structure approach*, in "IEEE Transactions on Automatic Control", October 2019, https://hal.inria.fr/hal-01966708

[10] D. EFIMOV, R. USHIROBIRA, J. A. MORENO, W. PERRUQUETTI.*Homogeneous Lyapunov functions: from converse design to numerical implementation*, in "SIAM Journal on Control and Optimization", December 2018, https://hal.inria.fr/hal-01851761

[11] N. ESPITIA, A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*Boundary time-varying feedbacks for fixed-time stabilization of constant-parameter reaction-diffusion systems*, in "Automatica", 2018, https://hal.inria.fr/hal-01962662

[12] H. ETHABET, D. RABEHI, D. EFIMOV, T. RAÏSSI.*Interval estimation for continuous-time switched linear systems*, in "Automatica", May 2018 [*DOI :* 10.1016/J.AUTOMATICA.2017.12.035], https://hal.inria.fr/hal-01888866

[13] M. FLIESS, C. JOIN, C. VOYANT.*Prediction bands for solar energy: New short-term time series forecasting techniques*, in "Solar Energy", May 2018, vol. 166, p. 519-528 [*DOI :* 10.1016/J.SOLENER.2018.03.049], https://hal-polytechnique.archives-ouvertes.fr/hal-01736518

[14] Z. KADER, C. FITER, L. HETEL, L. BELKOURA.*Stabilization by a relay control using non-quadratic Lyapunov functions*, in "Automatica", November 2018, vol. 97, p. 353 - 366 [*DOI :* 10.1016/J.AUTOMATICA.2018.08.020], https://hal.inria.fr/hal-01871693

[15] Z. KADER, G. ZHENG, J.-P. BARBOT.*Finite-time and asymptotic left inversion of nonlinear time-delay systems*, in "Automatica", April 2018, vol. 95, p. 283-292 [*DOI :* 10.1016/J.AUTOMATICA.2018.05.002], https://hal.archives-ouvertes.fr/hal-01781561

[16] T. KHARKOVSKAIA, D. EFIMOV, A. POLYAKOV, J.-P. RICHARD.*Design of interval observers and controls for PDEs using finite-element approximations*, in "Automatica", July 2018, vol. 93, p. 302 - 310 [*DOI :* 10.1016/J.AUTOMATICA.2018.03.016], https://hal.inria.fr/hal-01759948

[17] F. LOPEZ-RAMIREZ, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*Fixed-Time Output Stabilization and Fixed-Time Estimation of a Chain of Integrators*, in "International Journal of Robust and Nonlinear Control", December 2018, https://hal.inria.fr/hal-01817585

[18] F. LOPEZ-RAMIREZ, A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*Finite-Time and Fixed-Time Observer Design: Implicit Lyapunov Function Approach*, in "Automatica", January 2018, vol. 87, p. 52 - 60 [*DOI :* 10.1016/J.AUTOMATICA.2017.09.007], https://hal.inria.fr/hal-01622356

[19] M. MBOUP, C. JOIN, M. FLIESS, Y. WANG, G. ZHENG, D. EFIMOV, W. PERRUQUETTI.*Comments on "Differentiator application in altitude control for an indoor blimp robot"*, in "International Journal of Control", 2019 [*DOI :* 10.1080/00207179.2018.1500040], https://hal.inria.fr/hal-01832874

[20] R. ORTEGA, A. PYRKIN, A. BOBTSOV, D. EFIMOV, S. ARANOVSKIY.*A Globally Convergent Adaptive Indirect Field-Oriented Torque Controller for Induction Motors*, in "Asian Journal of Control", December 2018, vol. 00, p. 1 - 11 [*DOI :* 10.1002/ASJC.0000], https://hal.inria.fr/hal-01807720

[21] A. POLYAKOV, J.-M. CORON, L. ROSIER.*On Homogeneous Finite-Time Control for Linear Evolution Equation in Hilbert Space*, in "IEEE Transactions on Automatic Control", January 2018, vol. 63, n^o 9, p. 3143 - 3150 [*DOI :* 10.1109/TAC.2018.2797838], https://hal.inria.fr/hal-01695475

[22] A. POLYAKOV, D. EFIMOV, B. BROGLIATO.*Consistent Discretization of Finite-time and Fixed-time Stable Systems*, in "SIAM Journal on Control and Optimization", 2019, vol. 57, n^o 1, p. 78-103 [*DOI :* 10.1137/18M1197345], https://hal.inria.fr/hal-01838712

[23] A. POLYAKOV.*Sliding Mode Control Design Using Canonical Homogeneous Norm*, in "International Journal of Robust and Nonlinear Control", 2018, p. 1-29, https://hal.inria.fr/hal-01697797

[24] T. RAÏSSI, D. EFIMOV.*Some recent results on interval observers design of uncertain systems*, in "Automatisierungstechnik", March 2018, https://hal.inria.fr/hal-01888446

[25] J. G. RUEDA-ESCOBEDO, R. USHIROBIRA, D. EFIMOV, J. MORENO.*Gramian-based uniform convergent observer for stable LTV systems with delayed measurements*, in "International Journal of Control", 2018, https://hal.inria.fr/hal-01889193

[26] H. RÍOS, D. EFIMOV, J. A. MORENO, W. PERRUQUETTI, J. G. RUEDA-ESCOBEDO.*Time-Varying Parameter Identification Algorithms: Finite and Fixed-Time Convergence*, in "IEEE Transactions on Automatic Control", February 2018, https://hal.inria.fr/hal-01470957

[27] H. RÍOS, D. EFIMOV, W. PERRUQUETTI.*An Adaptive Sliding-Mode Observer for a Class of Uncertain Nonlinear Systems*, in "International Journal of Adaptive Control and Signal Processing", March 2018, vol. 32, n^{o} 3, p. 511 - 527 [*DOI :* 10.1002/ACS.2857], https://hal.inria.fr/hal-01889143

[28] H. RÍOS, L. HETEL, D. EFIMOV.*Robust output-feedback control for uncertain linear sampled-data systems: A 2D impulsive system approach*, in "Nonlinear Analysis: Hybrid Systems", May 2019, vol. 32, p. 177-201, https://hal.archives-ouvertes.fr/hal-01954885

[29] H. SILM, R. USHIROBIRA, D. EFIMOV, J.-P. RICHARD, W. MICHIELS.*A note on distributed finite-time observers*, in "IEEE Transactions on Automatic Control", May 2018 [*DOI :* 10.1109/TAC.2018.2838042], https://hal.inria.fr/hal-01827041

[30] Y. WANG, G. ZHENG, D. EFIMOV, W. PERRUQUETTI.*Differentiator application in altitude control for an indoor blimp robot*, in "International Journal of Control", March 2018, vol. 91, n^{o} 9, p. 2121 - 2130 [*DOI :* 10.1080/00207179.2018.1441549], https://hal.inria.fr/hal-01889244

[31] G. ZHENG, A. POLYAKOV, A. LEVANT.*Delay estimation via sliding mode for nonlinear time-delay systems*, in "Automatica", 2018, vol. 89, p. 266-273 [*DOI :* 10.1016/J.AUTOMATICA.2017.11.033], https://hal.inria.fr/hal-01649546

[32] K. ZIMENKO, A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*On simple scheme of finite/fixed-time control design*, in "International Journal of Control", 2018, https://hal.inria.fr/hal-01872533

[33] K. ZIMENKO, A. POLYAKOV, D. EFIMOV.*On Finite-Time Robust Stabilization via Nonlinear State Feedback*, in "International Journal of Robust and Nonlinear Control", 2018, https://hal.inria.fr/hal-01830102

Articles in National Peer-Reviewed Journal

[34] M. FLIESS, C. JOIN.*Two competing improvements of PID controllers: A comparison*, in "Automatique Control", January 2018, vol. 1, n^{o} 1, https://arxiv.org/abs/1801.10399 - 23 pages [*DOI :* 10.21494/ISTE.OP.2018.0284], https://hal-polytechnique.archives-ouvertes.fr/hal-01687952

Invited Conferences

[35] A. POLYAKOV, D. EFIMOV, B. BROGLIATO.*Consistent Discretization of Finite-time Stable Homogeneous Systems*, in "VSS 2018 - 15th International Workshop on Variable Structure Systems and Sliding Mode Control", Graz, Austria, IEEE, July 2018, p. 360-365 [*DOI :* 10.1109/VSS.2018.8460306], https://hal.inria.fr/hal-01794386

International Conferences with Proceedings

[36] P.-A. BLIMAN, D. EFIMOV, R. USHIROBIRA.*A class of nonlinear adaptive observers for SIR epidemic model*, in "ECC 2018 - European Control Conference", Limassol, Cyprus, June 2018, 6, https://hal.inria.fr/hal-01724989

[37] K. H. DEGUE, D. EFIMOV, J. LE NY, E. FÉRON.*Interval Observers for Secure Estimation in Cyber-Physical Systems*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Fontainebleau (FL), United States, December 2018, https://hal.inria.fr/hal-01888558

[38] D. EFIMOV, N. BARABANOV, R. ORTEGA.*Robust stability under relaxed persistent excitation conditions*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Fontainebleau (FL), United States, December 2018, https://hal.inria.fr/hal-01888575

[39] D. EFIMOV, E. FRIDMAN, W. PERRUQUETTI, J.-P. RICHARD.*On hyper-exponential output-feedback stabilization of a double integrator by using artificial delay*, in "Proc. 17th European Control Conference (ECC)", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01755086

[40] D. EFIMOV, A. POLYAKOV, A. Y. ALEKSANDROV.*On State-dependent Discretization of Stable Homogeneous Systems*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Fontainebleau (FL), United States, December 2018, https://hal.inria.fr/hal-01888564

[41] D. EFIMOV, J. SCHIFFER.*A new criterion for boundedness of solutions for a class of periodic systems*, in "Proc. 17th European Control Conference (ECC)", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01755084

[42] N. ESPITIA, A. POLYAKOV, D. EFIMOV, W. PERRUQUETTI.*On continuous boundary time-varying feedbacks for fixed-time stabilization of coupled reaction-diffusion systems*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Miami Beach, United States, December 2018, https://hal.inria.fr/hal-01892409

[43] R. FRANCO, H. RÍOS, D. EFIMOV, W. PERRUQUETTI.*Adaptive Estimation for Uncertain Nonlinear Systems: A Sliding-Mode Observer Approach*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Fontainebleau (FL), United States, December 2018, https://hal.inria.fr/hal-01888534

[44] T. KHARKOVSKAIA, D. EFIMOV, E. FRIDMAN, A. POLYAKOV, J.-P. RICHARD.*Interval estimation for second-order delay differential equations with delayed measurements and uncertainties*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Fontainebleau (FL), United States, December 2018, https://hal.inria.fr/hal-01888545

[45] F. LOPEZ-RAMIREZ, D. EFIMOV, A. POLYAKOV, W. PERRUQUETTI.*On Implicit Finite-Time and Fixed-Time ISS Lyapunov Functions*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Miami Beach, United States, December 2018, https://hal.inria.fr/hal-01888526

[46] F. Lopez-Ramirez, D. Efimov, A. Polyakov, W. Perruquetti. *On Necessary and Sufficient Conditions for Fixed-Time Stability of Continuous Autonomous Systems*, in "Proc. 17th European Control Conference (ECC)", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01755079

[47] M. Nancel, S. Aranovskiy, R. Ushirobira, D. Efimov, S. Poulmane, N. Roussel, G. Casiez. *Next-Point Prediction for Direct Touch Using Finite-Time Derivative Estimation*, in "Proceedings of the ACM Symposium on User Interface Software and Technology (UIST 2018)", Berlin, Germany, October 2018 [*DOI :* 10.1145/3242587.3242646], https://hal.inria.fr/hal-01893310

[48] G. Perozzi, D. Efimov, J.-M. Biannic, L. Planckaert, P. Coton. *Wind estimation algorithm for quadrotors using detailed aerodynamic coefficients*, in "ACC 2018 - American Control Conference", Milwaukee, WI, United States, June 2018, https://hal.inria.fr/hal-01703084

[49] J. G. Rueda-Escobedo, R. Ushirobira, D. Efimov, J. Moreno. *A Gramian-based observer with uniform convergence rate for delayed measurements*, in "ECC 2018 - 16th annual European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01895389

[50] T. Sanchez, A. Polyakov, L. Hetel, E. Fridman. *A switching controller for a class of MIMO bilinear time-delay systems*, in "2018 ICSEE - International Conference on the Science of Electrical Engineering", Eliat, Israel, December 2018, https://hal.inria.fr/hal-01919529

[51] T. Sanchez, A. Polyakov, J.-P. Richard, D. Efimov. *A robust Sliding Mode Controller for a class of bilinear delayed systems*, in "VSS18 - 15th International Workshop on Variable Structure Systems and Sliding Mode Control", Graz, Austria, April 2018, https://hal.inria.fr/hal-01763373

[52] J. Schiffer, D. Efimov, R. Ortega. *Almost Global Synchronization in Radial Multi-Machine Power Systems*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Fontainebleau (FL), United States, December 2018, https://hal.inria.fr/hal-01888553

[53] H. Silm, R. Ushirobira, D. Efimov, W. Michiels, J.-P. Richard, E. Fridman. *Comparison of the Time-Delay Margin of a Distributed and Centralized Observer*, in "ECC 2018 - European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01761390

[54] K. Zimenko, D. Efimov, A. Polyakov, A. Kremlev. *Control of Systems with Arbitrary Bounded Input Delay Using Implicit Lyapunov Function Technique*, in "Proc. 17th European Control Conference (ECC)", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01755078

[55] K. Zimenko, A. Polyakov, D. Efimov, W. Perruquetti. *Finite-time and fixed-time stabilization for integrator chain of arbitrary order*, in "ECC 2018 - 17th European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01797154

[56] K. Zimenko, A. Polyakov, D. Efimov, W. Perruquetti. *Generalized Feedback Homogenization and Stabilization of Linear MIMO Systems**, in "ECC 2018 - 17th European Control Conference", Limassol, Cyprus, June 2018, https://hal.inria.fr/hal-01797150

[57] K. Zimenko, A. Polyakov, D. Efimov. *On dynamical feedback control design for generalized homogeneous differential inclusions*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Miami Beach, United States, December 2018, https://hal.inria.fr/hal-01872548

Conferences without Proceedings

[58] D. DILEEP, W. MICHIELS, L. HETEL, J.-P. RICHARD.*Design of robust structurally constrained controllers for MIMO plants with time-delays*, in "European Control Conference 2018", Limassol, Cyprus, June 2018, https://hal.archives-ouvertes.fr/hal-01761889

[59] E. LEURENT, Y. BLANCO, D. EFIMOV, O.-A. MAILLARD.*Approximate Robust Control of Uncertain Dynamical Systems*, in "32nd Conference on Neural Information Processing Systems (NeurIPS 2018) Workshop", Montréal, Canada, December 2018, https://hal.archives-ouvertes.fr/hal-01931744

[60] J. THOMAS, C. FITER, L. HETEL, N. VAN DE WOUW, J.-P. RICHARD.*Dissipativity based stability criterion for aperiodic sampled-data systems subject to time-delay*, in "5th IFAC Conference on Analysis and Control of Chaotic Systems (CHAOS)", Eindhoven, Netherlands, October 2018, https://hal.archives-ouvertes.fr/hal-01893534

[61] J. THOMAS, L. HETEL, C. FITER, N. VAN DE WOUW, J.-P. RICHARD.*L2-stability criterion for systems with decentralized asynchronous controllers*, in "57th IEEE Conference on Decision and Control", Miami Beach, Florida, United States, December 2018, https://hal.archives-ouvertes.fr/hal-01893547

Research Reports

[62] M. FLIESS, C. JOIN.*Dynamic compensation and homeostasis: a feedback control perspective*, Ecole polytechnique X, January 2018, p. 1-22, https://arxiv.org/abs/1801.04959 , https://hal-polytechnique.archives-ouvertes.fr/hal-01684422

Scientific Popularization

[63] G. PEROZZI, D. EFIMOV, J.-M. BIANNIC, L. PLANCKAERT.*Trajectory Tracking For A Quadrotor Under Wind Perturbations: Sliding Mode Control With State-Dependent Gains*, in "Journal of The Franklin Institute", 2018 [*DOI :* 10.1016/J.JFRANKLIN.2018.04.042], https://hal.inria.fr/hal-01798700

Other Publications

[64] D. EFIMOV, J. SCHIFFER.*On boundedness of solutions of periodic systems: a multivariable cell structure approach*, September 2018, working paper or preprint, https://hal.inria.fr/hal-01755088

[65] G. PEROZZI.*A toolbox for quadrotors: from aerodynamic science to control theory*, January 2018, working paper or preprint, https://hal.inria.fr/hal-01696344

[66] G. PEROZZI.*CW-Quad Toolbox (User's guide)*, February 2018, working paper or preprint, https://hal.inria.fr/hal-01696347

[67] A. POLYAKOV.*On Optimal Control Problems for Generalized Homogeneous Systems*, February 2018, working paper or preprint, https://hal.inria.fr/hal-01705345

Project-Team RAPSODI

Reliable numerical approximations of dissipative systems.

IN COLLABORATION WITH: Laboratoire Paul Painlevé (LPP)

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

RESEARCH CENTER
Lille - Nord Europe

THEME
Numerical schemes and simulations

Table of contents

1. Team, Visitors, External Collaborators **309**
2. Overall Objectives **310**
2.1. Overall Objectives 310
2.2. Scientific Context 310
3. Research Program **311**
3.1. Design and analysis of structure-preserving schemes 311
3.1.1. Numerical analysis of nonlinear numerical methods 311
3.1.2. Design and analysis of asymptotic-preserving schemes 312
3.1.3. Design and stability analysis of numerical methods for low-Mach models 312
3.2. Optimizing the computational efficiency 312
3.2.1. High-order nonlinear numerical methods 312
3.2.2. A posteriori error control 313
3.2.3. Efficient computation of pairwise interactions in large systems of particles 313
4. Application Domains **313**
4.1. Porous media flows 313
4.2. Corrosion and concrete carbonation 313
4.3. Complex fluid flows 314
4.4. Stratigraphy 314
4.5. Low-frequency electromagnetism 315
5. Highlights of the Year **315**
6. New Software and Platforms **315**
7. New Results **316**
7.1. Numerical simulation of concrete carbonation 316
7.2. Modeling and numerical simulation of complex fluids 316
7.3. Stratigraphic modeling and simulation 317
7.4. Numerical simulation in low-frequency electromagnetism 317
7.5. Asymptotic analysis 317
7.6. Structure-preserving numerical methods 318
7.7. Cost reduction for numerical methods 319
7.8. Applied calculus of variations 320
7.9. Approximation theory 320
8. Bilateral Contracts and Grants with Industry **321**
9. Partnerships and Cooperations **321**
9.1. National Initiatives 321
9.1.1. ANR 321
9.1.2. LabEx CEMPI 321
9.2. International Research Visitors 322
9.2.1. Visits of International Scientists 322
9.2.2. Internships 322
9.2.3. Visits to International Teams 322
9.2.4. Research Stays Abroad 322
10. Dissemination **322**
10.1. Promoting Scientific Activities 322
10.1.1. Scientific Events Organization 322
10.1.1.1. General Chair, Scientific Chair 322
10.1.1.2. Member of the Organizing Committees 323
10.1.2. Journal 323
10.1.2.1. Member of the Editorial Boards 323
10.1.2.2. Reviewer - Reviewing Activities 323

10.1.3. Invited Talks 323
10.1.4. Research Administration 324
10.2. Teaching - Supervision - Juries 324
10.2.1. Teaching 324
10.2.2. Supervision 324
10.2.3. Juries 325
10.3. Popularization 325
11. Bibliography **325**

Project-Team RAPSODI

Creation of the Team: 2015 August 01, updated into Project-Team: 2017 November 01

Keywords:

Computer Science and Digital Science:

A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.1. - Numerical analysis of PDE and ODE

Other Research Topics and Application Domains:

B3. - Environment and planet
B3.3. - Geosciences
B3.3.1. - Earth and subsoil
B3.4. - Risks
B3.4.2. - Industrial risks and waste
B4. - Energy
B4.2. - Nuclear Energy Production
B4.2.1. - Fission

1. Team, Visitors, External Collaborators

Research Scientists

Clément Cancès [Inria, Researcher, HDR]
Maxime Herda [Inria, Researcher, from Oct. 2018]
Simon Lemaire [Inria, Researcher]

Faculty Members

Caterina Calgaro Zotto [Université de Lille, Associate Professor]
Claire Chainais-Hillairet [Team leader, Université de Lille, Professor, HDR]
Emmanuel Creusé [Université de Lille, Professor, until Aug. 2018, HDR]
Ingrid Lacroix-Violet [Université de Lille, Associate Professor, HDR]
Benoît Merlet [Université de Lille, Professor, HDR]
Thomas Rey [Université de Lille, Associate Professor]

External Collaborator

Emmanuel Creusé [Université Polytechnique Hauts-de-France, Professor, from Sep. 2018, HDR]

PhD Students

Ahmed Ait Hammou Oulhaj [Inria, until Feb. 2018]
Benoît Gaudeul [École Normale Supérieure de Rennes, from Sep. 2018]
Claire Colin-Lecerf [Université de Lille]
Nicolas Peton [IFPEn, until Oct. 2018]
Antoine Zurek [Université de Lille]

Post-Doctoral Fellows

Ahmed Ait Hammou Oulhaj [Université de Lille, from Mar. 2018 until Aug. 2018]
Florent Chave [Inria, from Dec. 2018]
David Maltese [Inria, until Sep. 2018]

Visiting Scientists

Jürgen Fuhrmann [WIAS Berlin, from May 2018 until Jun. 2018]
Naoufel Staili [Faculté des Sciences de Meknès, until Apr. 2018]
Juliette Venel [Université Polytechnique Hauts-de-France, until Jul. 2018]

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 project-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) [55], low-frequency electromagnetism [74], and mechanics of complex inhomogeneous fluids arising in avalanches [67] or in porous media [57].

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 apart, but combining them still leads to unsolved problems of crucial interest.

Let us mention for example the review paper by J. Droniou [83], 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 [83]. 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 [69]. The idea and the techniques introduced in [69] 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 [96], 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 [64] 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 [66] or [67]. 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 [79], 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 [76], the dead-oil model for porous media flows [70] or the Maxwell equations in their quasi-static approximation for some eddy current problems [74] and [75]. 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 of the recent breakthrough proposed by C. Cancès & C. Guichard [69], [4] 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 [83] 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 [71], [101]).

Recently, C. Chainais-Hillairet and co-authors [64], [72] and [73] 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 [64], [72] and [73]. The schemes proposed in [69] and [4] 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* [69] with the methodology of [64], [72], [73] 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 low-Mach models

We aim at extending the range of the `NS2DDV-M` software by introducing new physical models, like for instance the low-Mach model, which gives intermediate solutions between the compressible Navier–Stokes model and the incompressible Navier–Stokes one. This model was introduced in [99] as a limiting system which describes combustion processes at low Mach number in a confined region. Within this scope, we will propose a theoretical study for proving the existence of weak solutions for a particular class of models for which the dynamic viscosity of the fluid is a specific function of the density. We will propose also the extension of a combined Finite Volume-Finite Element method, initially developed for the simulation of incompressible and variable density flows, to this class of models.

3.2. Optimizing the computational efficiency

3.2.1. High-order nonlinear numerical methods

The numerical experiments carried out in [69] 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) [87].

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 schemes, that appear as an alternative to Finite Volume methods. Residual Distribution 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 [102], [95], with high-order accuracy [49], [48], or for parabolic problems [46], [47] make them very competitive. Relying on these breakthroughs, we aim at designing new Residual Distribution 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 [54], *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. [74] and [75]) 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 [66], [67] 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 [61], [105]. In the Finite Volume context, recent references deal with unsteady convection-diffusion equations [104], [52], [81] and [70]. 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 [90]. This reduces the $O(N^2)$ naive 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 [57], [56] 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 carbonation

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 [51]) as the modeling of corrosion in an underground repository (DPCM model developed by Bataillon *et al.* [55]) 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.

Variable-density, low-Mach flows have been widely studied in the recent literature because of their applicability in various phenomena such as flows in high-temperature gas reactors, meteorological flows, flows with convective and/or conductive heat transfer or combustion processes. In such cases, the resolution of the full compressible Navier–Stokes system is not adapted, because of the sound waves speed. The Boussinesq incompressible model is not a better alternative for such low-speed phenomena, because the compressibility effects can not be totally cancelled due to large variations of temperature and density. Consequently, some models have been formally derived, leading to the filtering of the acoustic waves by the use of some formal asymptotic expansions and two families of methods have been developed in the literature in order to compute these flows. We are interested in particular in the so-called pressure-based methods which are more robust than density-based solvers, although their range of validity is in general more limited.

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 *a posteriori* error estimators, is consequently necessary to reach a high precision of calculation in order to provide estimations as reliable as possible.

5. Highlights of the Year

5.1. Highlights of the Year

In 2018, the RAPSODI project-team was strongly involved in the organization of scientific events. In particular, in the framework of the LabEx CEMPI thematic semester on Numerical Analysis and PDEs, the following events were organized by RAPSODI members:

- the Mathematics-Enterprises Study Week, co-organized at LILLIAD Learning Center by E. Creusé from January 29 to February 2;
- the third edition of the ABPDE conference (on Asymptotic Behavior of systems of PDEs arising in physics and biology), co-organized at LILLIAD Learning Center by C. Cancès, C. Chainais-Hillairet, I. Lacroix-Violet, and T. Rey on August 28-31;
- the second edition of the One-day conference on Calculus of Variations, co-organized at Laboratoire Paul Painlevé by I. Lacroix-Violet and B. Merlet on October 12;
- the fifth edition of the Lille days on Numerical Analysis (dedicated to domain decomposition and its applications to PDEs), co-organized at Laboratoire Paul Painlevé by C. Calgaro Zotto and E. Creusé on November 13-14.

A research school on Mathematics for Nuclear Energy was also co-organized at the Roscoff Marine Station by C. Cancès on July 2-6, in partnership with the GdR MaNu. Let us as well mention the organization in the CANUM (national NUMerical Analysis Congress) at Cap d'Agde from May 28 to June 1 of three mini-symposia by members of the team: one by C. Cancès on cross-diffusion systems, one by S. Lemaire on polytopal discretization methods, and one co-organized by T. Rey on kinetic models. Team contributions finally include the co-organization by E. Creusé of the Maths Jobs Forum that was held in Paris on December 13, and the co-organization by A. Zurek of the Young Mathematicians Regional Tournament that was held in Laboratoire Paul Painlevé on April 14-15.

6. New Software and Platforms

6.1. Platform NS2DDV-M

`NS2DDV-M` is a Matlab code, developed by C. Calgaro Zotto, E. Creusé, and A. Mouton (CNRS research engineer at Université de Lille), for the simulation of homogeneous and inhomogeneous fluid flows by a combined Finite Volume-Finite Element method. The code is freely distributed, to allow for easy comparisons with concurrent codes on benchmark test-cases, and to promote new collaborations in the domain.
In 2018, a new version (v. 2.0) has been released, which contains a detailed documentation as well as some new functionalities, such as some post-processing tools and parallel computation capabilities.

7. New Results

7.1. Numerical simulation of concrete carbonation

In [20], C. Chainais-Hillairet, B. Merlet, and A. Zurek introduce and study a Finite Volume scheme for a concrete carbonation model proposed by Aiki and Muntean in [50]. This model consists in a system of two weakly coupled parabolic equations in a varying domain whose length is governed by an ordinary differential equation. The numerical scheme is obtained by a Euler discretization in time and a Scharfetter–Gummel discretization in space. The convergence of the scheme is established and the existence of a solution to the model is obtained as a by-product. Finally, some numerical experiments are performed to show the efficiency of the scheme.

In [45], A. Zurek studies the long-time regime of the moving interface appearing in the concrete carbonation model. He proves that the approximate free boundary, given by an implicit-in-time Finite Volume scheme, increases in time following a $\sqrt{t}$-law. This result is illustrated by numerical experiments.

7.2. Modeling and numerical simulation of complex fluids

In the context of C. Colin-Lecerf's PhD, C. Calgaro Zotto, C. Colin-Lecerf, and E. Creusé derive in [35] a combined Finite Volume-Finite Element scheme for a low-Mach model, in which a temperature field obeying an energy law is taken into account. The continuity equation is solved, whereas the state equation linking temperature, density, and thermodynamic pressure is imposed implicitly. Since the velocity field is not divergence-free, the projection method solving the momentum equation has to be adapted. This combined scheme preserves some steady states, and ensures a discrete maximum principle on the density. Numerical results are provided and compared to other approaches using purely Finite Element schemes, on a benchmark consisting in particular in a transient injection flow [58], [89], [53], as well as in the natural convection of a flow in a cavity [97], [93], [89], [53].

The theoretical study of the low-Mach limit system is a vast subject that has been considered by many authors. In particular, in [86], Embid establishes the local-in-time existence of classical solutions in Sobolev spaces. In [77], Danchin and Liao study the well-posedness issue in the critical Besov spaces, locally and globally, assuming that the initial density is close to a constant and that the initial velocity is small enough. Levermore *et al.* [98] consider the so-called ghost effect system, which is quite similar to the low-Mach system with thermal stress term added to the right-hand-side of the momentum equation, and they prove the local well-posedness of classical solutions for the Cauchy problem. In [94], Huang and Tan prove a local well-posedness result for strong solutions and also the existence and uniqueness of a global strong solution for the two-dimensional case. In [14], C. Calgaro Zotto, C. Colin-Lecerf, E. Creusé *et al.* investigate a specific low-Mach model for which the dynamic viscosity of the fluid is a specific function of the density. The model is reformulated in terms of the temperature and velocity, with nonlinear temperature equation, and strong solutions are considered. In addition to a local-in-time existence result for strong solutions, some convergence rates of the error between the approximation and the exact solution are obtained, following the same approach as Guillén-González *et al.* [91], [92].

Diffuse interface models, such as the Kazhikhov–Smagulov model, allow to describe some phase transition phenomena. In [15], C. Calgaro Zotto and co-workers investigate theoretically the combined Finite Volume-Finite Element scheme. They construct a fully discrete numerical scheme for approximating the two-dimensional Kazhikhov–Smagulov model, using a first-order time discretization and a splitting in time to allow the construction of the combined scheme. Consequently, at each time step, one only needs to solve two decoupled problems, the first one for the density (using the Finite Volume method) and the second one for the velocity and pressure (using the Finite Element method). The authors prove the stability of the combined scheme and the convergence towards the global-in-time weak solution of the model.

In [27], I. Lacroix-Violet *et al.* 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 difficulties 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.

In [34], I. Lacroix-Violet *et al.* generalize to the Navier–Stokes–Korteweg (with density-dependent viscosities satisfying the BD relation) and Euler–Korteweg systems a recent relative entropy proposed in [65]. As a concrete application, this helps justifying mathematically the convergence between global weak solutions of the quantum Navier–Stokes system and dissipative solutions of the quantum Euler system when the viscosity coefficient tends to zero. The results are based on the fact that Euler–Korteweg systems and corresponding Navier–Stokes–Korteweg systems can be reformulated through an augmented system. As a by-product of the analysis, Lacroix-Violet *et al.* show that this augmented formulation helps to define relative entropy estimates for the Euler–Korteweg systems in a simpler way and with less hypotheses compared to recent works [82], [88].

7.3. Stratigraphic modeling and simulation

Stratigraphy is a discipline of physics that aims at predicting the geological composition of the subsoil. In [44], N. Peton, C. Cancès *et al.* propose a new water flow driven forward stratigraphic model with the following particularities. First, the water surface flow is modelled at the continuous level, in opposition to what is currently done in this community. Second, the model incorporates a constraint on the erosion rate. A stable numerical scheme is proposed to simulate the model.

7.4. Numerical simulation in low-frequency electromagnetism

In [24], [28], E. Creusé and co-workers investigate the behavior of some Finite Element error estimators in the context of low-frequency elecromagnetism simulations, to underline the main differences in some practical situations. In addition, a more theoretical contribution is developed in [23], to prove the equivalence of some usual discrete gauge conditions. Once again, their numerical behaviors are compared on some characteristic benchmarks.

7.5. Asymptotic analysis

In [33], C. Cancès and co-workers derive the porous medium equation as the hydrodynamic limit of an interacting particle system which belongs to the family of exclusion processes with nearest neighbor exchanges. The main outcome of this work is to allow regions with vanishing density, where the dynamics turns out to degenerate. The convergence builds on a generalization of the entropy method and on suitable regularization of the dynamics.

In [29], A. Ait Hammou Oulhaj, C. Cancès, C. Chainais-Hillairet *et al.* study analytically and numerically the large time behavior of the solutions to a two-phase extension of the porous medium equation, which models the so-called seawater intrusion problem. They identify the self-similar solutions that correspond to steady states of a rescaled version of the problem. They finally provide numerical illustrations of the stationary states and exhibit numerical convergence rates.

In [13], C. Chainais-Hillairet *et al.* propose a new proof of existence of a solution to the scheme introduced in [63] which does not require any assumption on the time step. The result relies on the application of a topological degree argument which is based on the positivity and on uniform-in-time upper bounds of the approximate densities. They also establish uniform-in-time lower bounds satisfied by the approximate densities. These uniform-in-time upper and lower bounds ensure the exponential decay of the scheme towards the thermal equilibrium as shown in [63].

In [38], C. Chainais-Hillairet and M. Herda study the large-time behavior of solutions to Finite Volume discretizations of convection-diffusion equations or systems endowed with non-homogeneous Dirichlet and Neumann type boundary conditions. Their results concern various linear and nonlinear models such as Fokker–Planck equations, porous media equations, or drift-diffusion systems for semiconductors. For all of these models, some relative entropy principle is satisfied and implies exponential decay to the stationary state. They show that in the framework of Finite Volume schemes on orthogonal meshes, a large class of two-point monotone fluxes preserve this exponential decay of the discrete solution to the discrete steady state of the scheme.

In [32], M. Herda, T. Rey *et al.* are interested in the asymptotic analysis of a Finite Volume scheme for one-dimensional linear kinetic equations, with either Fokker–Planck or linearized BGK collision operator. Thanks to appropriate uniform estimates, they establish that the proposed scheme is asymptotic-preserving in the diffusive limit. Moreover, they adapt to the discrete framework the hypocoercivity method proposed by [80] to prove the exponential return to equilibrium of the approximate solution. They obtain decay estimates that are uniform in the diffusive limit. Finally, they present an efficient implementation of the proposed numerical schemes, and perform numerous numerical simulations assessing their accuracy and efficiency in capturing the correct asymptotic behaviors of the models.

In [26], M. Herda *et al.* consider various sets of Vlasov–Fokker–Planck equations modeling the dynamics of charged particles in a plasma under the effect of a strong magnetic field. For each of them, in a regime where the strength of the magnetic field is effectively stronger than that of collisions, they first formally derive asymptotically reduced models. In this regime, strong anisotropic phenomena occur; while equilibrium along magnetic field lines is asymptotically reached the asymptotic models capture a nontrivial dynamics in the perpendicular directions. They do check that in any case the obtained asymptotic model defines a well-posed dynamical system and when self-consistent electric fields are neglected they provide a rigorous mathematical justification of the formally derived systems. In this last step they provide a complete control on solutions by developing anisotropic hypocoercive estimates.

7.6. Structure-preserving numerical methods

The design and the analysis of numerical methods preserving at the discrete level the key features of the continuous models is one of the core tasks of the RAPSODI project-team. C. Cancès was invited to write a review paper [16] on energy stable numerical methods for complex porous media flows. The paper addresses three different approaches: monotonicity-based numerical methods like two-point flux approximation Finite Volumes, as well as two methods based on multi-point flow approximation that are either based on upwinding or on positive local dissipation tensors.

Concerning methods based on upwinding, A. Ait Hammou Oulhaj, C. Cancès, and C. Chainais-Hillairet extend in [12] the nonlinear Control Volume Finite Element scheme of [69] to the discretization of Richards equation modeling unsaturated flows in porous media. This strategy is also applied in [30] by A. Ait Hammou Oulhaj and D. Maltese for the simulation of seawater intrusion in the subsoil nearby coastal regions. The scheme proposed in [30] is still convergent if the porous medium is anisotropic, in opposition to the energy-diminishing scheme analyzed in [11] by A. Ait Hammou Oulhaj, which is designed to be accurate in the long-time regime studied in [29]. Besides, an implicit Euler-Finite Volume scheme for a degenerate cross-diffusion system describing the ion transport through biological membranes is analyzed in [17] by C. Cancès, C. Chainais-Hillairet *et al*. The strongly coupled equations for the ion concentrations include drift terms involving the electric potential, which is coupled to the concentrations through the Poisson equation. The cross-diffusion system possesses a formal gradient flow structure revealing nonstandard degeneracies, which lead to considerable mathematical difficulties. The Finite Volume scheme is based on two-point flux approximations with “double” upwind mobilities. It preserves the structure of the continuous model like non-negativity, upper bounds, and entropy dissipation.

Concerning methods based on positive local dissipation tensors, C. Cancès, C. Chainais-Hillairet *et al.* propose in [18] a nonlinear Discrete Duality Finite Volume scheme to approximate the solutions of drift diffusion equations. The scheme is built to preserve at the discrete level even on severely distorted meshes the energy/energy dissipation relation. In [37], C. Cancès and co-workers propose a Finite Element scheme for the numerical approximation of degenerate parabolic problems in the form of a nonlinear anisotropic Fokker–Planck equation. The scheme is energy-stable, only involves physically motivated quantities in its definition, and is able to handle general unstructured grids. Its convergence is rigorously proven thanks to compactness arguments, under very general assumptions. Although the scheme is based on Lagrange Finite Elements of degree 1, it is locally conservative after a local post-processing giving rise to an equilibrated flux. This also allows to derive a guaranteed *a posteriori* error estimate for the approximate solution. Numerical experiments are presented in order to give evidence of a very good behavior of the proposed scheme in various situations involving strong anisotropy and drift terms.

C. Cancès *et al.* derive in [36] a model of degenerate Cahn–Hilliard type for the phase segregation in incompressible multiphase flows. The model is obtained as the Wasserstein gradient flow of a Ginzburg–Landau energy with the constraint that the sum of the volume fractions must stay equal to 1. The resulting model differs from the classical degenerate Cahn–Hilliard model (see [106], [85]) and is closely related to a model proposed by E and collaborators [84], [100]. Besides the derivation of the model, the convergence of a minimizing movement scheme is proven in [36]. The Wasserstein gradient flow structure of the PDE system governing multiphase flows in porous media has recently been highlighted in [68]. The model can thus be approximated by means of the minimizing movement (or JKO) scheme, that C. Cancès *et al.* solve in [19] thanks to the ALG2-JKO scheme proposed in [60]. The numerical results are compared to a classical upstream mobility Finite Volume scheme, for which strong stability properties can be established.

In [42], S. Lemaire builds a bridge between the Hybrid High-Order [78] and Virtual Element [59] methods, which are the two main new-generation approaches to the arbitrary-order approximation of PDEs on meshes with general, polytopal cells. The Virtual Element method writes in functional terms and is naturally conforming; at the opposite, the Hybrid High-Order method writes in algebraic terms and is naturally nonconforming. It has been remarked a few years ago that the Hybrid High-Order method can be viewed as a nonconforming version of the Virtual Element method. In [42], S. Lemaire ends up unifying the Hybrid High-Order and Virtual Element approaches by showing that the Virtual Element method can be reformulated as a (newborn) conforming Hybrid High-Order method. This parallel has interesting consequences: it allows important simplifications in the *a priori* analysis of Virtual Element methods, and sheds new light on the differences between conforming and nonconforming Virtual Element methods, in particular in terms of mesh assumptions.

In [31], I. Lacroix-Violet *et al.* are interested in the numerical integration in time of nonlinear Schrödinger equations using different methods preserving the energy or a discrete analog of it. In particular, they give a rigorous proof of the order of the relaxation method (presented in [62] for cubic nonlinearities) and they propose a generalized version that allows to deal with general power law nonlinearities. Numerical simulations for different physical models show the efficiency of these methods.

7.7. Cost reduction for numerical methods

In [22], S. Lemaire *et al.* design and analyze (in the periodic setting) nonconforming multiscale methods for highly oscillatory elliptic problems, which are applicable on coarse grids that may feature general polytopal cells. Two types of methods are introduced: a Finite Element-type method, that generalizes classical nonconforming multiscale Finite Element methods to general meshes and to arbitrary-order polynomial cell boundary conditions, and a Virtual Element-type method, that allows, up to the computation of an adequate projection, to compute less oscillatory basis functions for equivalent precision. The Virtual Element-type method is based on the Hybrid High-Order framework [78]. As standard with such multiscale approaches, the general workflow of the method splits into an offline, massively parallelizable stage, where all fine-scale computations are performed, and the online, fully-coarse-scale stage.

In [25], T. Rey *et al.* extend the Fast Kinetic Scheme (FKS) 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 conservative 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 dynamics.

In [43], T. Rey *et al.* present high-order, fully explicit time integrators for nonlinear collisional kinetic equations, including the full Boltzmann equation. The methods, called projective integration, first take a few small steps with a simple, explicit method (forward Euler) to damp out the stiff components of the solution. Then, the time derivative is estimated and used in a Runge–Kutta method of arbitrary order. The procedure can be recursively repeated on a hierarchy of projective levels to construct telescopic projective integration methods. We illustrate the method with numerical results in one and two spatial dimensions.

7.8. Applied calculus of variations

In [41], B. Merlet *et al.* study a variational problem which models the behavior of topological singularities on the surface of a biological membrane in P_β-phase (see [103]). The problem combines features of the Ginzburg–Landau model in 2D and of the Mumford–Shah functional. As in the classical Ginzburg–Landau theory, a prescribed number of point vortices appear in the moderate energy regime; the model allows for discontinuities, and the energy penalizes their length. The novel phenomenon here is that the vortices have a fractional degree $1/m$ with m prescribed. Those vortices must be connected by line discontinuities to form clusters of total integer degrees. The vortices and line discontinuities are therefore coupled through a topological constraint. As in the Ginzburg–Landau model, the energy is parameterized by a small length scale $\varepsilon > 0$. B. Merlet *et al.* perform a complete Γ-convergence analysis of the model as $\varepsilon \downarrow 0$ in the moderate energy regime. Then, they study the structure of minimizers of the limit problem. In particular, the line discontinuities of a minimizer solve a variant of the Steiner problem.

In [21], B. Merlet *et al.* consider a generalization of branched transport in arbitrary dimension and codimension: minimize the h-mass of some oriented k-dimensional branched surface in $\mathbf{R}^n$ with some prescribed boundary. Attached to the surface is a multiplicity $m(x)$ which is not necessarily an integer and is a conserved quantity (Kirchhoff current law is satisfied at branched points). The h-mass is defined as the integral of a cost $h(|m(x)|)$ over the branched surface. As usual in branched transportation, the cost function is a lower-semicontious, sublinear increasing function with $h(0) = 0$ (for instance $h(m) = \sqrt{1 + am^2}$ if $m \neq 0$ and $h(0) = 0$). For numerical purpose, it is convenient to approximate the measure defined by the k-dimensional surfaces by smooth functions in $\mathbf{R}^n$. In this spirit, B. Merlet *et al.* propose phase field approximations of the branched surfaces and of their energy in the spirit of the Ambrosio–Tortorelli functional. The convergence of these approximations towards the original k-dimensional branched transport problem is established in [21] in the sense of Γ-convergence. Next, considering the cost $h(m) = \sqrt{1 + am^2}$ and sending a to 0, a phase field approximation of the Plateau problem is obtained. Numerical experiments show the efficiency of the method. These numerical results are exceptional as they are obtained without any guess on the topology of the minimizing k-surface (as opposed to methods based on parameterizations of the k-surface). In [39], B. Merlet *et al.* establish new results on the approximation of k-dimensional surfaces (k-rectifiable currents) by polyhedral surfaces with convergence in h-mass and with preservation of the boundary (the approximating polyhedral surface has the same boundary as the limit). This approximation result is required in the convergence study of [21].

7.9. Approximation theory

In [40], M. Herda *et al.* propose a new iterative algorithm for the calculation of sum of squares decompositions of polynomials, reformulated as positive interpolation. The method is based on the definition of a dual functional G from values at interpolation points. The domain of G, the boundary of the domain and the behavior of G at infinity are analyzed in details. In the general case, G is closed convex. For univariate polynomials in the context of the Lukacs representation, G is coercive and strictly convex which yields a unique critical point, corresponding to a sum of squares decomposition of G. Various descent algorithms are evoked. Numerical examples are provided, for univariate and bivariate polynomials.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

C. Cancès supervised the PhD thesis of N. Peton at IFPEn from October 15, 2015 to October 12, 2018. The bilateral contract enters the framework-agreement between Inria and IFPEn.

9. Partnerships and Cooperations

9.1. National Initiatives

9.1.1. ANR

C. Chainais-Hillairet is a member of the ANR MOONRISE project. 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: MOdels, Oscillations, and NumeRIcal SchEmes

Type: Fondements du numérique (DS0705) - 2014

ANR reference: ANR-14-CE23-0007

Coordinator: F. Méhats (Université de Rennes 1)

Duration: October 2014 - June 2019

C. Chainais-Hillairet and T. Rey are members of the ANR MOHYCON project. The MOHYCON project is related to the analysis and simulation of multiscale models of semiconductors. As almost all current electronic technology involves the use of semiconductors, there is a strong interest for modeling and simulating the behavior of such devices, which was recently reinforced by the development of organic semiconductors used for example in solar panels or in mobile phones and television screens (among others).

Title: multiscale MOdels and HYbrid numerical methods for semiCONductors

Type: Société de l'information et de la communication (DS07) - 2017

ANR reference: ANR-17-CE40-0027

Coordinator: M. Bessemoulin-Chatard (CNRS and Université de Nantes)

Duration: January 2018 - December 2020

9.1.2. LabEx CEMPI

Title: Centre Européen pour les Mathématiques, la Physique et leurs Interactions

Coordinator: S. De Bièvre (Université de Lille)

Duration: January 2012 - December 2019

Partners: Laboratoire Paul Painlevé and Laser Physics department (PhLAM), Université de Lille

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.

9.2. International Research Visitors

9.2.1. Visits of International Scientists

The RAPSODI project-team invited several scientists in 2018. The following people came for long visits:

- J. Venel (Université Polytechnique Hauts-de-France) visited Inria Lille until July;
- J. Fuhrmann (WIAS Berlin) was invited for 1 month between May and June, thanks to a support of the LabEx CEMPI;
- A. Vasseur (UT Austin) was invited for 1 month in June, also thanks to a support of the LabEx CEMPI.

The following people came for shorter visits:

- S. Krell (Université de Nice) came in Lille on February 12-15;
- M. Rodrigues (Université de Rennes 1) came in Lille on December 3-7;
- M. Breden (TU Munich) came on December 17-21.

9.2.2. Internships

N. Staili, PhD student in the Faculté des Sciences de Meknès, came in Lille for a three-month visit between January and April.

9.2.3. Visits to International Teams

C. Cancès visited D. Matthes (TU Munich) during 1 week on December 3-7 to collaborate on the variational derivation of multiphase flow models.

9.2.4. Research Stays Abroad

A. Zurek spent 2 months (October-November) in the research team of A. Jüngel at TU Vienna to collaborate on the numerical simulation of a biofilm model. He was supported by the Institut Français d'Autriche and by EKINOX CNRS grant (Laboratoire Paul Painlevé).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organization

10.1.1.1. General Chair, Scientific Chair

Four scientific events were organized by RAPSODI members in the framework of the LabEx CEMPI thematic semester on Numerical Analysis and PDEs:

- the Mathematics-Enterprises Study Week, co-organized at LILLIAD Learning Center by E. Creusé from January 29 to February 2;
- the third edition of the ABPDE conference (on Asymptotic Behavior of systems of PDEs arising in physics and biology), co-organized at LILLIAD Learning Center by C. Cancès, C. Chainais-Hillairet, I. Lacroix-Violet, and T. Rey on August 28-31;
- the second edition of the One-day conference on Calculus of Variations, co-organized at Laboratoire Paul Painlevé by I. Lacroix-Violet and B. Merlet on October 12;
- the fifth edition of the Lille days on Numerical Analysis (dedicated to domain decomposition and its applications to PDEs), co-organized at Laboratoire Paul Painlevé by C. Calgaro Zotto and E. Creusé on November 13-14.

C. Cancès co-organized a research school on Mathematics for Nuclear Energy at the Roscoff Marine Station on July 2-6, in partnership with the GdR MaNu. C. Chainais-Hillairet was part of the scientific board for this event.

In the CANUM (national NUMerical Analysis Congress) at Cap d'Agde from May 28 to June 1, three mini-symposia were organized by members of the team: one by C. Cancès on cross-diffusion systems, one by S. Lemaire on polytopal discretization methods, and one co-organized by T. Rey on kinetic models.

E. Creusé co-organized the Maths Jobs Forum that was held in Paris on December 13.

A. Zurek co-organized the Young Mathematicians Regional Tournament that was held in Laboratoire Paul Painlevé on April 14-15.

10.1.1.2. Member of the Organizing Committees

The whole team RAPSODI was involved in the organization of the ABPDE III conference.

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 and of the International Journal on Finite Volumes.

10.1.2.2. Reviewer - Reviewing Activities

RAPSODI team members are regular reviewers for all the main international journals in numerical analysis and PDEs.

10.1.3. Invited Talks

C. Cancès was an invited speaker in the International workshop on PDEs, optimal transport, and applications, held on October 17-20 in Essaouira. He was also one of the speakers in the mini-symposium on polytopal discretization methods organized by S. Lemaire at CANUM (Cap d'Agde, May 28-June 1). He finally gave several seminars in Amiens, Orsay, Strasbourg, and Munich.

C. Chainais-Hillairet was invited to give a talk in a mini-symposium on finite volume methods at the fifteenth edition of the International Conference Zaragoza-Pau on Mathematics and its Applications held in Jaca on September 10-12. She also gave a seminar in Lyon (ICJ).

B. Gaudeul presented a poster at the AMaSiS (Applied Mathematics and Simulation for Semiconductors) conference held in Berlin on October 8-10.

M. Herda gave a talk in the same conference in Berlin, and another one for the Day of the Nord-Pas-de-Calais Mathematics Research Federation on October 3 in Lille. He was also one of the speakers in the mini-symposium on kinetic models co-organized by T. Rey at CANUM (Cap d'Agde, May 28-June 1).

I. Lacroix-Violet gave several seminars in Marseille (I2M), Dijon (IMB), Montpellier (IMAG), and Rennes.

S. Lemaire was invited to give a talk in a mini-symposium on polytopal discretization methods in the biennial Congress of the Italian Society of Applied and Industrial Mathematics (SIMAI) held in Rome on July 2-6. He also gave two seminars in Montpellier (IMAG), and Inria Paris, and gave a talk for the ANEDP team day at the Laboratoire Paul Painlevé.

D. Maltese gave a talk at CANUM (Cap d'Agde, May 28-June 1).

B. Merlet was an invited speaker in the congress Geometric Measure Theory in Verona held on June 11-15, and in the Workshop in Calculus of Variations organized at Paris-Diderot on June 25-27.

T. Rey was an invited speaker for the MAFRAN Days (Mathematical Frontiers in the Analysis of Many-particle Systems) held in Cambridge on September 24-26, and in the mini-workshop Innovative Trends in the Numerical Analysis and Simulation of Kinetic Equations organized in Oberwolfach on December 16-22. He also gave several seminars in Montpellier (IMAG), Paris-Dauphine (CEREMADE), Imperial College London, and Paris-Descartes.

A. Zurek was invited to give a talk in a special session on the mathematical problems arising from materials and biological science in the twelfth edition of the AIMS Conference on Dynamical Systems, Differential Equations, and Applications held in Taipei on July 5-9. He also presented a poster at CANUM (Cap d'Agde, May 28-June 1).

10.1.4. Research Administration

C. Cancès is the head of the MaNu research group (GdR MaNu) funded by the Institute for Mathematical Sciences and its Interactions (INSMI) of the French National Center for Research (CNRS).

E. Creusé has in charge to develop some actions promoted by AMIES (Agency for the Mathematics in Interaction with the Enterprise and the Society). More particularly, his action in 2018 was devoted to several characteristic points: management of some PEPS (First Support for Exploratory Projects), discussions to initiate collaborations between academic researchers in mathematics and industrial partners, participation to the monthly AMIES meeting, organization of the Mathematics-Enterprises Study Week in Lille in January as well as of the Maths Jobs Forum in Paris in December.

I. Lacroix-Violet, B. Merlet, and T. Rey are elected members of the Conseil du Laboratoire Paul Painlevé. I. Lacroix-Violet is also a member of the Jury de domaine.

T. Rey is in charge of the organization of the weekly seminar of the ANEDP team of the Laboratoire Paul Painlevé. He is also a member of the team of the Opération Postes.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

RAPSODI team members are strongly involved in teaching at the Université de Lille. C. Calgaro Zotto is in charge of the Master of Mathematical Engineering. B. Merlet is in charge of the Master 2 of Scientific Computing. E. Creusé was responsible of the “Cursus Master en Ingénierie Mathématiques” until August. He is since September director of the Mathematics Department of the Université Polytechnique Hauts-de-France. C. Cancès gave lectures at Centrale Lille. S. Lemaire gave lectures in the Master 2 of Scientific Computing.

C. Chainais-Hillairet gave 16h of courses in the international Summer School “The way to become a mathematician” in the Harbin Institute of Technology in July.

10.2.2. Supervision

Post-doc: F. Chave arrived in December to work on high-order polytopal discretization methods for electromagnetism; supervisors: E. Creusé and S. Lemaire.

PhD: L. Ferrari defended his PhD thesis on “Phase field approximations for branched transport problems” on October 5; advisors: A. Chambolle (CNRS & CMAP, École Polytechnique) and B. Merlet.

PhD: N. Peton defended his PhD thesis on “Numerical methods for a stratigraphic model with nonlinear diffusion and moving frontier areas” on October 12; advisors: C. Cancès, Q.-H. Tran (IFPEn), and S. Wolf (IFPEn).

PhD in progress: C. Colin-Lecerf, on the “Analyse numérique et simulations de modèles multi-fluides”, since 10/01/2015; advisors: C. Calgaro Zotto and E. Creusé.

PhD in progress: A. Zurek, on the “Numerical and theoretical analysis of models describing the corrosion of materials”, since 10/01/2016; advisors: C. Chainais-Hillairet and B. Merlet.

PhD in progress: B. Gaudeul, on the “Numerical approximation of cross-diffusion systems arising in physics and biology”, since 09/01/2018; advisors: C. Cancès and C. Chainais-Hillairet.

Master internship: A. El Keurti, on the “Study of upwind finite volume schemes for nonlocal transport”, from November 2017 to July 2018; advisor: T. Rey.

Master internship: A. Latrech, on the “Numerical study of local minimizers of a Bose–Einstein energy functional in 2D”, from January to July; advisors: G. Dujardin (Inria Lille) and I. Lacroix-Violet.

Master internship: T. Ebrahimipourfaez, on “Entropy-diminishing finite volume schemes for cross-diffusion systems”, from April to July; advisor: C. Chainais-Hillairet.

Master internship: B. Gaudeul, on “Numerical schemes for a Nernst–Planck–Poisson model”, from April to July; advisor: C. Chainais-Hillairet.

Master internship: C. Marinel, on the “Implementation and parallelization of a finite element code with *a posteriori* error estimators”, from June to August; advisor: E. Creusé.

Master internship in progress: N. Aghouzzaf, on the “Design of high-order numerical time integrators for kinetic equations”, since December; advisor: T. Rey.

10.2.3. Juries

E. Creusé reported on O. Gorynina’s PhD thesis, defended on February 22, 2018 at Université Bourgogne-Franche Comté. Title: Éléments finis adaptatifs pour l’équation des ondes instationnaire.

B. Merlet was a member of the jury of the PhD thesis of A. Julia, defended on October 9, 2018 at Sorbonne Paris Cité. Title: Functions with bounded variations on a current.

C. Calgaro Zotto and C. Cancès are members of the Jury de l’Agrégation de Mathématiques, which is a national hiring committee for the highest level of high-school teachers.

C. Cancès was part of the selection committee for an associate professor (MCF) position at the Laboratoire de Mathématiques d’Orsay.

C. Chainais-Hillairet was part of the selection committee for a full professor (PR) position at Aix-Marseille Université.

10.3. Popularization

C. Calgaro Zotto is in charge of the communication of the Laboratoire Paul Painlevé. She organizes various events which promote mathematics among young people:

- les “Mathématiques itinérantes”;
- la “Semaine des mathématiques”;
- la collection “Stages scientifiques en Seconde”.

Members of the team participate regularly to these actions.

S. Lemaire gave a talk at Inria Lille for the internal scientific popularization event “30 minutes of Science” in June.

11. Bibliography

Major publications by the team in recent years

[1] 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", 2017, vol. 25, n$^{\text{o}}$ 3, p. 147-168 [*DOI :* 10.1515/JNMA-2016-0007], https://hal.archives-ouvertes.fr/hal-01250709

[2] C. CALGARO, E. CREUSÉ, T. GOUDON, S. KRELL.*Simulations of non homogeneous viscous flows with incompressibility constraints*, in "Mathematics and Computers in Simulation", 2017, vol. 137, p. 201-225, https://hal.archives-ouvertes.fr/hal-01246070

[3] C. CANCÈS, T. GALLOUËT, L. MONSAINGEON.*Incompressible immiscible multiphase flows in porous media: a variational approach*, in "Analysis & PDE", 2017, vol. 10, n$^{\text{o}}$ 8, p. 1845–1876 [*DOI :* 10.2140/APDE.2017.10.1845], https://hal.archives-ouvertes.fr/hal-01345438

[4] C. CANCÈS, C. GUICHARD.*Numerical analysis of a robust free energy diminishing Finite Volume scheme for parabolic equations with gradient structure*, in "Foundations of Computational Mathematics", 2017, vol. 17, n° 6, p. 1525-1584, https://hal.archives-ouvertes.fr/hal-01119735

[5] C. CHAINAIS-HILLAIRET, B. MERLET, A. VASSEUR.*Positive Lower Bound for the Numerical Solution of a Convection-Diffusion Equation*, in "FVCA8 2017 - International Conference on Finite Volumes for Complex Applications VIII", Lille, France, Springer, June 2017, p. 331-339 [*DOI :* 10.1007/978-3-319-57397-7_26], https://hal.archives-ouvertes.fr/hal-01596076

[6] D. A. DI PIETRO, A. ERN, S. LEMAIRE.*An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators*, in "Computational Methods in Applied Mathematics", June 2014, vol. 14, n° 4, p. 461-472 [*DOI :* 10.1515/CMAM-2014-0018], https://hal.archives-ouvertes.fr/hal-00978198

[7] G. DIMARCO, R. LOUBÈRE, J. NARSKI, T. REY.*An efficient numerical method for solving the Boltzmann equation in multidimensions*, in "Journal of Computational Physics", 2018, vol. 353, p. 46-81 [*DOI :* 10.1016/J.JCP.2017.10.010], https://hal.archives-ouvertes.fr/hal-01357112

[8] F. FILBET, M. HERDA.*A finite volume scheme for boundary-driven convection-diffusion equations with relative entropy structure*, in "Numerische Mathematik", 2017, vol. 137, n° 3, p. 535-577, https://hal.archives-ouvertes.fr/hal-01326029

[9] I. LACROIX-VIOLET, A. VASSEUR.*Global weak solutions to the compressible quantum Navier–Stokes equation and its semi-classical limit*, in "Journal de Mathématiques Pures et Appliquées", 2018, vol. 114, p. 191-210, https://hal.archives-ouvertes.fr/hal-01347943

[10] B. MERLET.*A highly anisotropic nonlinear elasticity model for vesicles I. Eulerian formulation, rigidity estimates and vanishing energy limit*, in "Arch. Ration. Mech. Anal.", 2015, vol. 217, n° 2, p. 651–680 [*DOI :* 10.1007/S00205-014-0839-5], https://hal.archives-ouvertes.fr/hal-00848547

Publications of the year

Articles in International Peer-Reviewed Journal

[11] A. AIT HAMMOU OULHAJ.*Numerical analysis of a finite volume scheme for a seawater intrusion model with cross-diffusion in an unconfined aquifer*, in "Numerical Methods for Partial Differential Equations", May 2018 [*DOI :* 10.1002/NUM.22234], https://hal.archives-ouvertes.fr/hal-01432197

[12] 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*, in "ESAIM: Mathematical Modelling and Numerical Analysis", 2018, vol. 52, n° 4, p. 1532-1567 [*DOI :* 10.1051/M2AN/2017012], https://hal.archives-ouvertes.fr/hal-01372954

[13] M. BESSEMOULIN-CHATARD, C. CHAINAIS-HILLAIRET.*Uniform-in-time Bounds for approximate Solutions of the drift-diffusion System*, in "Numerische Mathematik", 2018, https://hal.archives-ouvertes.fr/hal-01659418

[14] C. CALGARO, C. COLIN, E. CREUSÉ, E. ZAHROUNI.*Approximation by an iterative method of a low Mach model with temperature dependent viscosity*, in "Mathematical Methods in the Applied Sciences", 2018 [*DOI :* 10.1002/MMA.5342], https://hal.archives-ouvertes.fr/hal-01801242

[15] C. CALGARO, M. EZZOUG, E. ZAHROUNI.*Stability and convergence of an hybrid finite volume-finite element method for a multiphasic incompressible fluid model*, in "Communications on Pure and Applied Analysis", March 2018, vol. 17, n^{o} 2, p. 429-448, https://hal.archives-ouvertes.fr/hal-01586201

[16] C. CANCÈS.*Energy stable numerical methods for porous media flow type problems*, in "Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles", 2018, vol. 73, 78 [*DOI :* 10.2516/OGST/2018067], https://hal.archives-ouvertes.fr/hal-01953395

[17] C. CANCÈS, C. CHAINAIS-HILLAIRET, A. GERSTENMAYER, A. JÜNGEL.*Convergence of a Finite-Volume Scheme for a Degenerate Cross-Diffusion Model for Ion Transport*, in "Numerical Methods for Partial Differential Equations", 2018, https://arxiv.org/abs/1801.09408 , https://hal.archives-ouvertes.fr/hal-01695129

[18] C. CANCÈS, C. CHAINAIS-HILLAIRET, S. KRELL.*Numerical analysis of a nonlinear free-energy diminishing Discrete Duality Finite Volume scheme for convection diffusion equations*, in "Computational Methods in Applied Mathematics", 2018, vol. 18, n^{o} 3, p. 407-432, https://arxiv.org/abs/1705.10558 - Special issue on "Advanced numerical methods: recent developments, analysis and application" [*DOI :* 10.1515/CMAM-2017-0043], https://hal.archives-ouvertes.fr/hal-01529143

[19] C. CANCÈS, T. GALLOUËT, M. LABORDE, L. MONSAINGEON.*Simulation of multiphase porous media flows with minimizing movement and finite volume schemes*, in "European Journal of Applied Mathematics", 2018 [*DOI :* 10.1017/S0956792518000633], https://hal.archives-ouvertes.fr/hal-01700952

[20] C. CHAINAIS-HILLAIRET, B. MERLET, A. ZUREK.*Convergence of a finite volume scheme for a parabolic system with a free boundary modeling concrete carbonation*, in "ESAIM: Mathematical Modelling and Numerical Analysis", June 2018, vol. 52, n^{o} 2, p. 457-480, https://hal.archives-ouvertes.fr/hal-01477543

[21] A. CHAMBOLLE, L. A. D. FERRARI, B. MERLET.*Variational approximation of size-mass energies for k-dimensional currents*, in "ESAIM: Control, Optimisation and Calculus of Variations", April 2018, https://arxiv.org/abs/1710.08808 , https://hal.archives-ouvertes.fr/hal-01622540

[22] M. CICUTTIN, A. ERN, S. LEMAIRE.*A Hybrid High-Order method for highly oscillatory elliptic problems*, in "Computational Methods in Applied Mathematics", 2018 [*DOI :* 10.1515/CMAM-2018-0013], https://hal.archives-ouvertes.fr/hal-01467434

[23] E. CREUSÉ, P. DULAR, S. NICAISE.*About the gauge conditions arising in Finite Element magnetostatic problems*, in "Computers and Mathematics with Applications", 2018, https://hal.archives-ouvertes.fr/hal-01955649

[24] E. CREUSÉ, Y. LE MENACH, S. NICAISE, F. PIRIOU, R. TITTARELLI.*Two Guaranteed Equilibrated Error Estimators for Harmonic Formulations in Eddy Current Problems*, in "Computers and Mathematics with Applications", 2018, https://hal.archives-ouvertes.fr/hal-01955692

[25] G. DIMARCO, R. LOUBÈRE, J. NARSKI, T. REY.*An efficient numerical method for solving the Boltzmann equation in multidimensions*, in "Journal of Computational Physics", 2018, vol. 353, p. 46-81 [*DOI :* 10.1016/J.JCP.2017.10.010], https://hal.archives-ouvertes.fr/hal-01357112

[26] M. HERDA, L. M. M. RODRIGUES.*Anisotropic Boltzmann-Gibbs dynamics of strongly magnetized Vlasov-Fokker-Planck equations*, in "Kinetic and Related Models ", 2018, https://arxiv.org/abs/1610.05138 , https://hal.archives-ouvertes.fr/hal-01382854

[27] I. LACROIX-VIOLET, A. VASSEUR.*Global weak solutions to the compressible quantum navier-stokes equation and its semi-classical limit*, in "Journal de Mathématiques Pures et Appliquées", 2018, vol. 114, p. 191-210, https://arxiv.org/abs/1607.06646 , https://hal.archives-ouvertes.fr/hal-01347943

[28] R. TITTARELLI, Y. LE MENACH, F. PIRIOU, E. CREUSÉ, S. NICAISE, J.-P. DUCREUX.*Comparison of Numerical Error Estimators for Eddy Current Problems solved by FEM*, in "IEEE Transactions on Magnetics", 2018, vol. 54, n^{o} 3, https://hal.archives-ouvertes.fr/hal-01645591

Other Publications

[29] A. AIT HAMMOU OULHAJ, C. CANCÈS, C. CHAINAIS-HILLAIRET, P. LAURENÇOT.*Large time behavior of a two phase extension of the porous medium equation*, March 2018, https://arxiv.org/abs/1803.10476 - 28 pages, 14 figures, https://hal.archives-ouvertes.fr/hal-01752759

[30] A. AIT HAMMOU OULHAJ, D. MALTESE.*Positive nonlinear Control Volume Finite Element scheme for an anisotropic seawater intrusion model with cross-diffusion in an unconfined aquifer*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01906872

[31] C. BESSE, S. DESCOMBES, G. DUJARDIN, I. LACROIX-VIOLET.*Energy preserving methods for nonlinear schrodinger equations*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01951527

[32] M. C. BESSEMOULIN-CHATARD, M. HERDA, T. REY.*Hypocoercivity and diffusion limit of a finite volume scheme for linear kinetic equations*, December 2018, 36 pages, 9 figures, 2 tables, https://hal.archives-ouvertes.fr/hal-01957832

[33] O. BLONDEL, C. CANCÈS, M. SASADA, M. SIMON.*Convergence of a Degenerate Microscopic Dynamics to the Porous Medium Equation*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01710628

[34] D. BRESCH, M. GISCLON, I. LACROIX-VIOLET.*On Navier-Stokes-Korteweg and Euler-Korteweg Systems: Application to Quantum Fluids Models*, March 2018, https://arxiv.org/abs/1703.09460 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01496960

[35] C. CALGARO, C. COLIN, E. CREUSÉ.*A combined Finite Volumes -Finite Elements method for a low-Mach model*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01574894

[36] C. CANCÈS, D. MATTHES, F. NABET.*A two-phase two-fluxes degenerate Cahn-Hilliard model as constrained Wasserstein gradient flow*, 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01665338

[37] C. CANCÈS, F. NABET, M. VOHRALÍK.*Convergence and a posteriori error analysis for energy-stable finite element approximations of degenerate parabolic equations*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01894884

[38] C. CHAINAIS-HILLAIRET, M. HERDA.*Large-time behavior of a family of finite volume schemes for boundary-driven convection-diffusion equations*, October 2018, https://arxiv.org/abs/1810.01087 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01885015

[39] A. CHAMBOLLE, L. A. D. FERRARI, B. MERLET.*Strong approximation in h-mass of rectifiable currents under homological constraint*, June 2018, https://arxiv.org/abs/1806.05046 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01813234

[40] B. DESPRÉS, M. HERDA.*Iterative Calculation of Sum Of Squares*, December 2018, https://arxiv.org/abs/1812.02444 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01946539

[41] M. GOLDMAN, B. MERLET, V. MILLOT.*A Ginzburg-Landau model with topologically induced free discontinuities*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01643795

[42] S. LEMAIRE.*Bridging the Hybrid High-Order and Virtual Element methods*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01902962

[43] W. MELIS, T. REY, G. SAMAEY.*Projective and telescopic projective integration for the nonlinear BGK and Boltzmann equations*, December 2018, https://arxiv.org/abs/1712.06362 - 35 pages, 2 annexes, 12 figures, https://hal.archives-ouvertes.fr/hal-01666346

[44] N. PETON, C. CANCÈS, D. GRANJEON, Q.-H. TRAN, S. WOLF.*Numerical scheme for a water flow-driven forward stratigraphic model*, September 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01870347

[45] A. S. F. ZUREK.*Numerical approximation of a concrete carbonation model: study of the $\sqrt{t}$-law of propagation*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01839277

References in notes

[46] 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

[47] 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

[48] 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

[49] 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

[50] T. AIKI, A. MUNTEAN.*Existence and uniqueness of solutions to a mathematical model predicting service life of concrete structure*, in "Adv. Math. Sci. Appl.", 2009, vol. 19, p. 109-129

[51] 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

[52] 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

[53] M. AVILA, J. PRINCIPE, R. CODINA.*A finite element dynamical nonlinear subscale approximation for the low Mach number flow equations*, in "J. Comput. Phys.", 2011, vol. 230, n^{o} 22, p. 7988–8009, http://dx.doi.org/10.1016/j.jcp.2011.06.032

[54] 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

[55] 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

[56] J. BEAR, Y. BACHMAT.*Introduction to modeling of transport phenomena in porous media*, Springer, 1990, vol. 4

[57] J. BEAR.*Dynamic of Fluids in Porous Media*, American Elsevier, New York, 1972

[58] A. BECCANTINI, E. STUDER, S. GOUNAND, J.-P. MAGNAUD, T. KLOCZKO, C. CORRE, S. KUDRIAKOV.*Numerical simulations of a transient injection flow at low Mach number regime*, in "Internat. J. Numer. Methods Engrg.", 2008, vol. 76, n^{o} 5, p. 662–696, http://dx.doi.org/10.1002/nme.2331

[59] L. BEIRÃO DA VEIGA, F. BREZZI, A. CANGIANI, G. MANZINI, L. D. MARINI, A. RUSSO.*Basic principles of virtual element methods*, in "Math. Models Methods Appl. Sci. (M3AS)", 2013, vol. 23, n^{o} 1, p. 199–214

[60] J.-D. BENAMOU, G. CARLIER, M. LABORDE.*An augmented Lagrangian approach to Wasserstein gradient flows and applications*, in "Gradient flows: from theory to application", ESAIM Proc. Surveys, EDP Sci., Les Ulis, 2016, vol. 54, p. 1–17, https://doi.org/10.1051/proc/201654001

[61] 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

[62] C. BESSE.*Analyse numérique des systèmes de Davey–Stewartson*, Université Bordeaux 1, 1998

[63] M. BESSEMOULIN-CHATARD, C. CHAINAIS-HILLAIRET.*Exponential decay of a finite volume scheme to the thermal equilibrium for drift-diffusion systems*, in "J. Numer. Math.", 2017, vol. 25, n^{o} 3, p. 147–168, https://doi.org/10.1515/jnma-2016-0007

[64] 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

[65] D. BRESCH, P. NOBLE, J.-P. VILA.*Relative entropy for compressible Navier-Stokes equations with density dependent viscosities and various applications*, in "LMLFN 2015—low velocity flows—application to low Mach and low Froude regimes", ESAIM Proc. Surveys, EDP Sci., Les Ulis, 2017, vol. 58, p. 40–57

[66] 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

[67] 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

[68] C. CANCÈS, T. O. GALLOUËT, L. MONSAINGEON.*Incompressible immiscible multiphase flows in porous media: a variational approach*, in "Anal. PDE", 2017, vol. 10, n^o^ 8, p. 1845–1876, https://doi.org/10.2140/apde.2017.10.1845

[69] 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

[70] 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

[71] 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

[72] 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

[73] 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

[74] 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

[75] 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

[76] 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

[77] R. DANCHIN, X. LIAO.*On the well-posedness of the full low Mach number limit system in general critical Besov spaces*, in "Commun. Contemp. Math.", 2012, vol. 14, n^o^ 3, 1250022, 47, https://doi.org/10.1142/S0219199712500228

[78] D. A. DI PIETRO, A. ERN, S. LEMAIRE.*An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators*, in "Comput. Methods Appl. Math.", 2014, vol. 14, n^o^ 4, p. 461–472, https://doi.org/10.1515/cmam-2014-0018

[79] 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)

[80] J. DOLBEAULT, C. MOUHOT, C. SCHMEISER.*Hypocoercivity for linear kinetic equations conserving mass*, in "Trans. Amer. Math. Soc.", 2015, vol. 367, n^{o} 6, p. 3807–3828, https://doi.org/10.1090/S0002-9947-2015-06012-7

[81] 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

[82] D. DONATELLI, E. FEIREISL, P. MARCATI.*Well/ill posedness for the Euler–Korteweg–Poisson system and related problems*, in "Comm. Partial Differential Equations", 2015, vol. 40, p. 1314-1335

[83] 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

[84] W. E, P. PALFFY-MUHORAY.*Phase separation in incompressible systems*, in "Phys. Rev. E", Apr 1997, vol. 55, p. R3844–R3846, https://link.aps.org/doi/10.1103/PhysRevE.55.R3844

[85] C. M. ELLIOTT, H. GARCKE.*On the Cahn-Hilliard equation with degenerate mobility*, in "SIAM J. Math. Anal.", 1996, vol. 27, n^{o} 2, p. 404–423, http://dx.doi.org/10.1137/S0036141094267662

[86] P. EMBID.*Well-posedness of the nonlinear equations for zero Mach number combustion*, in "Comm. Partial Differential Equations", 1987, vol. 12, n^{o} 11, p. 1227–1283, https://doi.org/10.1080/03605308708820526

[87] 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

[88] J. GIESSELMANN, C. LATTANZIO, A.-E. TZAVARAS.*Relative energy for the Korteweg theory and related Hamiltonian flows in gas dynamics*, in "Arch. Rational Mech. Analysis", 2017, vol. 223, p. 1427-1484

[89] V. GRAVEMEIER, W. A. WALL.*Residual-based variational multiscale methods for laminar, transitional and turbulent variable-density flow at low Mach number*, in "Internat. J. Numer. Methods Fluids", 2011, vol. 65, n^{o} 10, p. 1260–1278, http://dx.doi.org/10.1002/fld.2242

[90] 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

[91] F. GUILLÉN-GONZÁLEZ, P. DAMÁZIO, M. A. ROJAS-MEDAR.*Approximation by an iterative method for regular solutions for incompressible fluids with mass diffusion*, in "J. Math. Anal. Appl.", 2007, vol. 326, n^{o} 1, p. 468–487, http://dx.doi.org/10.1016/j.jmaa.2006.03.009

[92] F. GUILLÉN-GONZÁLEZ, M. SY.*Iterative method for mass diffusion model with density dependent viscosity*, in "Discrete Contin. Dyn. Syst. Ser. B", 2008, vol. 10, n^{o} 4, p. 823–841, http://dx.doi.org/10.3934/dcdsb.2008.10.823

[93] V. HEUVELINE.*On higher-order mixed FEM for low Mach number flows: application to a natural convection benchmark problem*, in "Internat. J. Numer. Methods Fluids", 2003, vol. 41, n^{o} 12, p. 1339–1356, http://dx.doi.org/10.1002/fld.454

[94] F. HUANG, W. TAN.*On the strong solution of the ghost effect system*, in "SIAM J. Math. Anal.", 2017, vol. 49, n^{o} 5, p. 3496–3526, https://doi.org/10.1137/16M106964X

[95] 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

[96] S. JIN.*Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations*, in "SIAM, J. Sci. Comput.", 1999, vol. 21, p. 441-454

[97] P. LE QUÉRÉ, R. MASSON, P. PERROT.*A Chebyshev collocation algorithm for 2D non-Boussinesq convection*, in "Journal of Computational Physics", 1992, vol. 103, p. 320-335

[98] S. LEVERMORE, W. SUN, K. TRIVISA.*Local well-posedness of a ghost effect system*, in "Indiana Univ. J.", 2011

[99] A. MAJDA, J. SETHIAN.*The derivation and numerical solution of the equations for zero Mach number combustion*, in "Combustion Science and Technology", 1985, vol. 42, p. 185–205

[100] F. OTTO, W. E.*Thermodynamically driven incompressible fluid mixtures*, in "J. Chem. Phys.", 1997, vol. 107, n^{o} 23, p. 10177-10184, https://doi.org/10.1063/1.474153

[101] 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

[102] 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

[103] D. RUPPEL, E. SACKMANN.*On defects in different phases of two-dimensional lipid bilayers*, in "J. Phys. France", 1983, vol. 44, n^{o} 9, p. 1025-1034 [*DOI :* 10.1051/JPHYS:019830044090102500], https://jphys.journaldephysique.org/articles/jphys/abs/1983/09/jphys_1983__44_9_1025_0/jphys_1983__44_9_1025_0.html

[104] 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

[105] 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

[106] P. G. DE GENNES.*Dynamics of fluctuations and spinodal decomposition in polymer blends*, in "J. Chem. Phys.", 1980, vol. 72, p. 4756-4763

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

1. Team, Visitors, External Collaborators **337**
2. Overall Objectives **338**
2.1. Introduction 338
2.2. Reengineering and remodularization 338
2.3. Constructs for modular and isolating programming languages 339
3. Research Program **339**
3.1. Software Reengineering 339
3.1.1. Tools for understanding applications 340
3.1.2. Remodularization analyses 340
3.1.3. Software Quality 341
3.2. Language Constructs for Modular Design 341
3.2.1. Traits-based program reuse 341
3.2.2. Reconciling Dynamic Languages and Isolation 342
4. Application Domains **343**
4.1. Programming Languages and Tools 343
4.2. Software Reengineering 343
5. Highlights of the Year **343**
5.1.1. Release of Pharo 7 343
5.1.2. Awards 343
6. New Software and Platforms **343**
6.1. Moose 343
6.2. Pharo 344
6.3. Pillar 344
7. New Results **344**
7.1. Dynamic Languages: Virtual Machines 344
7.2. Dynamic Languages: Language Constructs for Modular Design 345
7.3. Software Reengineering 346
7.4. Dynamic Languages: Debugging 346
7.5. Blockchain 347
8. Bilateral Contracts and Grants with Industry **348**
8.1. Bilateral Contracts with Industry 348
8.1.1. BlockChain 348
8.1.2. Pharo Consortium 348
8.2. Bilateral Grants with Industry 348
8.2.1. Thales CIFRE 348
8.2.2. Remodularization of Architecture 349
9. Partnerships and Cooperations **349**
9.1. Regional Initiatives 349
9.1.1. CAR IMT Douai 349
9.1.2. CPER DATA 349
9.2. National Initiatives 349
9.3. European Initiatives 349
9.4. International Initiatives 350
9.5. International Research Visitors 350
9.5.1. Visits of International Scientists 350
9.5.2. Visits to International Teams 351
10. Dissemination **351**
10.1. Promoting Scientific Activities 351
10.1.1. Scientific Events Organisation 351

10.1.2. Scientific Events Selection 351
10.1.2.1. Chair of Conference Program Committees 351
10.1.2.2. Reviewer 351
10.1.3. Journal 351
10.1.4. Scientific Expertise 352
10.2. Teaching - Supervision - Juries 352
10.2.1. Teaching 352
10.2.2. Supervision 353
10.3. Popularization 353
10.3.1. Articles and contents 353
10.3.2. Education 353
10.3.3. Interventions 353
11. Bibliography **353**

Project-Team RMOD

Creation of the Project-Team: 2009 July 01

Keywords:

Computer Science and Digital Science:

A2. - Software
A2.1. - Programming Languages
A2.1.3. - Object-oriented programming
A2.1.8. - Aspect-oriented programming
A2.1.10. - Domain-specific languages
A2.1.12. - Dynamic languages
A2.3.1. - Embedded systems
A2.5. - Software engineering
A2.5.1. - Software Architecture & Design
A2.5.3. - Empirical Software Engineering
A2.5.4. - Software Maintenance & Evolution
A2.6. - Infrastructure software
A2.6.3. - Virtual machines

Other Research Topics and Application Domains:

B2. - Health
B2.7. - Medical devices
B5. - Industry of the future
B5.9. - Industrial maintenance
B6.5. - Information systems
B7. - Transport and logistics

1. Team, Visitors, External Collaborators

Research Scientists

Stéphane Ducasse [Team leader, Inria, Senior Researcher, HDR]
Marcus Denker [Inria, Researcher]

Faculty Members

Nicolas Anquetil [Univ des sciences et technologies de Lille, Associate Professor, HDR]
Anne Etien [Univ de Lille, Associate Professor, HDR]
Damien Pollet [Univ de Lille, Associate Professor]
Vincent Aranega [Univ de Lille, Associate Professor, from Sep 2018]

External Collaborators

Guillermo Polito [CNRS]
Olivier Auverlot [Univ de Lille]
Esteban Lorenzano [Inria Soft]
Santiago Bragagnolo [InriaTech]

Technical Staff

Clément Béra [Inria, until Mar 2018]
Steven Costiou [Inria, from Oct 2018]

Cyril Ferlicot-Delbecque [Inria, from Oct 2018]
Pavel Krivanek [Inria]
Denis Kudriashov [Inria, until Apr 2018]
Guillaume Larcheveque [Inria, from May 2018, granted by ANR CARNOT Abondement project]
Allex Oliveira [Inria, from May 2018]
Christophe Demarey [Inria, Engineers, 70%]

PhD Students

Lionel Akue [Inria, from Nov 2018]
Julien Delplanque [Univ de Lille]
Thomas Dupriez [Ecole Normale Supérieure Paris, from Oct 2018]
Brice Govin [Thales, 3rd year, granted by CIFRE, until 2018]
Carolina Hernandez Phillips [Inria, from Oct 2018]
Jason Lecerf [CEA]
Matteo Marra [VUB Brussels]
Pablo Tesone [Ecole des Mines de Douai]

Post-Doctoral Fellow

Henrique Santos Camargos Rocha [Inria]

Visiting Scientists

Abdelghani Alidra [University of Skikda Algeria, from Nov 2018]
Andy Amoordon [Utocat, from Oct 2018]
Jan Bliznicenko [University of Prague , Mar 2018]
Abdelhakim Bouremel [University of Skikda Algeria, May 2018]
Christopher Fuhrman [Ecole de technologie superieure de montreal, from Oct 2018]
Tomohiro Oda [Key Technology Laboratory, Japan, from Aug 2018 until Sep 2018]
Giuseppe Antonio Pierro [University of Cagliari, from Aug 2018]
Ronie Salgado Faila [University of Chile at Santiago, Chile, from Aug 2018 until Sep 2018]
Serge Demeyer [Universiteit Antwerpen, Belgium, from Jun 2018 until Jul 2018]

Administrative Assistant

Julie Jonas [Inria]

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 [44]. 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 [41]. 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 [50]. 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.

3. Research Program

[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.

3.1. 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 [35], [34]. 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) [63] 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 [24]. 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 [51]. Until now, few works have attempted to identify layers in practice: Mudpie [65] is a first cut at identifying cycles between packages as well as package groups potentially representing layers. DSM (dependency structure matrix) [64], [59] 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 [47]. 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 [66], [38].

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 [61], [36] and classboxes [25] 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 [61], [36]. 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 [33], [55], [26], [37] and several type systems were defined [39], [62], [56], [49].

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 [28]. 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 [36], stateful [27], and freezable [37]) 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 [54]. 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 [32] then from Smalltalk [42].

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 [46]). 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 [29], as well as controlling the access to reflective features [30], [31] are important challenges. We plan to:

- Study the isolation abstractions available in erights (http://www.erights.org) [53], [52], and Java's class loader strategies [48], [43].
- Categorize the different reflective features of languages such as CLOS [45], Python and Smalltalk [57] and identify suitable isolation mechanisms and infrastructure [40].
- Assess different isolation models (access rights, capabilities [58]...) 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 [29],
  - the use of encapsulation policies as a basis to restrict the interfaces of the controlled objects [60],
  - 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...) [57].

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 25 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 analyzing software systems to support understanding and continuous development as well as software quality analysis.

5. Highlights of the Year

5.1. Highlights of the Year

5.1.1. Release of Pharo 7

We released a release candidate for Pharo 7, with a release to be expected early 2019. More information at http://pharo.org.

5.1.2. Awards

- Guillermo Polito, Pablo Tesone, Esteban Lorenzano and Nicolás Passerini won the 1st place in the Innovation Technologies Award at ESUG 2018.
- Christian Marlon Souza Couto, Henrique Rocha, and Ricardo Terra. A Quality-oriented Approach to Recommend Move Method Refactorings. 1st place in 17th Brazilian Symposium on Software Quality, SBQS p. 11—20, ACM, New York, NY, USA, 2018.

6. New Software and Platforms

6.1. Moose

Moose: Software and Data Analysis Platform
KEYWORDS: Software engineering - Meta model - Software visualisation
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. The development of Moose has been evaluated to 200 man/year.

Mots-cles : MetaModeling, Program Visualization, Software metrics, Code Duplication, Software analyses, Parsers

- Participants: Anne Etien, Nicolas Anquetil, Olivier Auverlot, Stéphane Ducasse, Julien Delplanque, Guillaume Larcheveque, Cyril Ferlicot-Delbecque and Pavel Krivanek
- Partners: Université de Berne - Sensus - Synectique - Pleiad - USI - Vrije Universiteit Brussel
- Contact: Stéphane Ducasse
- URL: http://www.moosetechnology.org

6.2. Pharo

KEYWORDS: Live programmation objet - Reflective system - Web Application
FUNCTIONAL DESCRIPTION: Pharo is a pure object reflective and dynamic language inspired by Smalltalk. In addition, Pharo comes with a full advanced programming environment developed under the MIT License. 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. Pharo 60 got 100 contributors world-wide. It is used by around 30 universities, 15 research groups and around 40 companies.

- Participants: Christophe Demarey, Clement Bera, Damien Pollet, Esteban Lorenzano, Marcus Denker, Stéphane Ducasse and Guillermo Polito
- Partners: BetaNine - Reveal - Inceptive - Netstyle - Feenk - ObjectProfile - GemTalk Systems - Greyc Université de Caen - Basse-Normandie - Université de Berne - Yesplan - RMod - Pleiad - Sensus - Université de Bretagne Occidentale - École des Mines de Douai - ENSTA - Uqbar foundation Argentina - LAM Research - ZWEIDENKER - LifeWare - JPMorgan Chase - KnowRoaming - ENIT - Spesenfuchs - FINWorks - Esug - FAST - Ingenieubüro Schmidt - Projector Software - HRWorks - Inspired.org - Palantir Solutions - High Octane - Soops - Osoco - Ta Mère SCRL - University of Yaounde 1 - Software Quality Laboratory, University of Novi Sad - Software Institute Università della Svizzera italiana - Universdad Nacional de Quilmes - UMMISCO IRD - Université technique de Prague
- Contact: Marcus Denker
- URL: http://www.pharo.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. It is used in the tools developed by Feenk.com.

- Partner: Feenk
- Contact: Stéphane Ducasse
- URL: https://github.com/Pillar-markup/pillar

7. New Results

7.1. Dynamic Languages: Virtual Machines

Assessing primitives performance on multi-stage execution. Virtual machines, besides the interpreter and just-in-time compiler optimization facilities, also include a set of primitive operations that the client language can use. Some of these are essential and cannot be performed in any other way. Others are optional: they can be expressed in the client language but are often implemented in the virtual machine to improve performance when the just-in-time compiler is unable to do so (start-up performance, speculative optimizations not implemented or not mature enough, etc.). In a hybrid runtime, where code is executed by an interpreter and a just-in-time compiler, the implementor can choose to implement optional primitives in the client language , in the virtual machine implementation language (typically C or C++), or on top of the just-in-time compiler back-end. This raises the question of the maintenance and performance trade-offs of the different alternatives. As a case study, we implemented the String comparison optional primitive in each case. The paper describes the different implementations, discusses the maintenance cost of each of them and evaluates for different string sizes the execution time in Cog, a Smalltalk virtual machine. [18]

Fully Reflective Execution Environments: Virtual Machines for More Flexible Software. VMs are complex pieces of software that implement programming language semantics in an efficient, portable, and secure way. Unfortunately, mainstream VMs provide applications with few mechanisms to alter execution semantics or memory management at run time. We argue that this limits the evolvability and maintainability of running systems for both, the application domain, e.g., to support unforeseen requirements, and the VM domain, e.g., to modify the organization of objects in memory. This work explores the idea of incorporating reflective capabilities into the VM domain and analyzes its impact in the context of software adaptation tasks. We characterize the notion of a fully reflective VM, a kind of VM that provides means for its own observability and modifiability at run time. This enables programming languages to adapt the underlying VM to changing requirements. We propose a reference architecture for such VMs and present TruffleMATE as a prototype for this architecture. We evaluate the mechanisms TruffleMATE provides to deal with unanticipated dynamic adaptation scenarios for security, optimization, and profiling aspects. In contrast to existing alternatives, we observe that TruffleMATE is able to handle all scenarios, using less than 50 lines of code for each, and without interfering with the application's logic. [2]

7.2. Dynamic Languages: Language Constructs for Modular Design

Dynamic Software Update from Development to Production. Dynamic Software Update (DSU) solutions update applications while they are executing. These solutions are typically used in production to minimize application downtime, or in integrated development environments to provide live programming support. Each of these scenarios presents different challenges, forcing existing solutions to be designed with only one of these use cases in mind. For example, DSUs for live programming typically do not implement safe point detection or instance migration, while production DSUs require manual generation of patches and lack IDE integration. Also, these solutions have limited ability to update themselves or the language core libraries, and some of them present execution penalties outside the update window. We propose a DSU (gDSU) that works for both live programming and production environments. Our solution implements safe update point detection using call stack manipulation and a reusable instance migration mechanism to minimize manual intervention in patch generation. Moreover, it also offers updates of core language libraries and the update mechanism itself. This is achieved by the incremental copy of the modified objects and an atomic commit operation. We show that our solution does not affect the global performance of the application and it presents only a run-time penalty during the update window. Our solution is able to apply an update impacting 100,000 instances in 1 second. In this 1 second, only during 250 milliseconds the application is not responsive. The rest of the time the application runs normally while gDSU is looking for the safe update point. The update only requires to copy the elements that are modified. [6]

Implementing Modular Class-based Reuse Mechanisms on Top of a Single Inheritance VM. Code reuse is a good strategy to avoid code duplication and speed up software development. Existing object-oriented programming languages propose different ways of combining existing and new code such as e.g., single inheritance, multiple inheritance, Traits or Mixins. All these mechanisms present advantages and disadvantages and there are situations that require the use of one over the other. To avoid the complexity of implementing a virtual machine (VM), many of these mechanisms are often implemented on top of an existing high-performance VM, originally meant to run a single inheritance object-oriented language. These implementations require thus a mapping between the programming model they propose and the execution model provided by the VM. Moreover, reuse mechanisms are not usually composable, nor it is easy to implement new ones for a given language. We propose a modular meta-level runtime architecture to implement and combine different code reuse mechanisms. This architecture supports dynamic combination of several mechanisms without affecting runtime performance in a single inheritance object-oriented VM. It includes moreover a reflective Meta-Object Protocol to query and modify classes using the programming logical model instead of the underlying low-level runtime model. Thanks to this architecture, we implemented Stateful Traits, Mixins, CLOS multiple inheritance, CLOS Standard Method Combinations and Beta prefixing in a modular and composable way. [15]

7.3. Software Reengineering

A Reflexive and Automated Approach to Syntactic Pattern Matching in Code Transformations. Empowering software engineers often requires to let them write code transformations. However existing automated or tool-supported approaches force developers to have a detailed knowledge of the internal representation of the underlying tool. While this knowledge is time consuming to master, the syntax of the language, on the other hand, is already well known to developers and can serve as a strong foundation for pattern matching. Pattern languages with metavariables (that is variables holding abstract syntax subtrees once the pattern has been matched) have been used to help programmers define program transformations at the language syntax level. The question raised is then the engineering cost of metavariable support. Our contribution is to show that, with a GLR parser, such patterns with metavariables can be supported by using a form of runtime reflexivity on the parser internal structures. This approach allows one to directly implement such patterns on any parser generated by a parser generation framework, without asking the pattern writer to learn the AST structure and node types. As a use case for that approach we describe the implementation built on top of the SmaCC (Smalltalk Compiler Compiler) GLR parser generator framework. This approach has been used in production for source code transformations on a large scale. We will express perspectives to adapt this approach to other types of parsing technologies. [12]

Relational Database Schema Evolution: An Industrial Case Study. Modern relational database management systems provide advanced features allowing, for example, to include behavior directly inside the database (stored procedures). These features raise new difficulties when a database needs to evolve (e.g. adding a new table). To get a better understanding of these difficulties, we recorded and studied the actions of a database architect during a complex evolution of the database at the core of a software system. From our analysis, problems faced by the database architect are extracted, generalized and explored through the prism of software engineering. Six problems are identified: (1) difficulty in analyzing and visualizing dependencies between database's entities, (2) difficulty in evaluating the impact of a modification on the database, (3) replicating the evolution of the database schema on other instances of the database, (4) difficulty in testing database's functionalities, (5) lack of synchronization between the IDE's internal model of the database and the database actual state and (6) absence of an integrated tool enabling the architect to search for dependencies between entities, generate a patch or access up to date PostgreSQL documentation. We suggest that techniques developed by the software engineering community could be adapted to help in the development and evolution of relational databases. [10]

A Quality-oriented Approach to Recommend Move Method Refactorings. Refactoring is an important activity to improve software internal structure. Even though there are many refactoring approaches, very few consider their impact on the software quality. We propose a software refactoring approach based on quality attributes. We rely on the measurements of the Quality Model for Object Oriented Design (QMOOD) to recommend Move Method refactorings that improve software quality. In a nutshell, given a software system S, our approach recommends a sequence of refactorings R1,R2,...,Rn that result in system versions S1, S2,..., Sn, where quality (Si+1) > quality (Si). We empirically calibrated our approach, using four systems, to find the best criteria to measure the quality improvement. We performed three types of evaluation to verify the usefulness of our implemented tool, named QMove. First, we applied our approach on 13 open-source systems achieving an average recall of 84.2%. Second, we compared QMove with two state-of-art refactoring tools (JMove and JDeodorant) on the 13 previously evaluated systems, and QMove showed better recall, precision, and f-score values than the others. Third, we evaluated QMove, JMove, and JDeodorant in a real scenario with two proprietary systems on the eyes of their software architects. As result, the experts positively evaluated a greater number of QMove recommendations. [14]

7.4. Dynamic Languages: Debugging

Collectors. Observing and modifying object-oriented programs often means interacting with objects. At runtime, it can be a complex task to identify those objects due to the live state of the program. Some objects may exist for only a very limited period of time, others can be hardly reachable because they are never stored in variables. To address this problem we present Collectors. They are dedicated objects which can collect objects of interest at runtime and present them to the developer. Collectors are non-intrusive, removable code instrumentations. They can be dynamically specified and injected at runtime. They expose an API to allow their specification and the access to the collected objects. We present an implementation of Collectors in Pharo, a Smalltalk dialect. We enrich the Pharo programming and debugging environment with tools that support the Collectors API. We illustrate the use of these API and tools through the collection and the logging of specific objects in a running IOT application. [9]

Rotten Green Tests: a First Analysis. Unit tests are a tenant of agile programming methodologies, and are widely used to improve code quality and prevent code regression. A passing (green) test is usually taken as a robust sign that the code under test is valid. However, we have noticed that some green tests contain assertions that are never executed; these tests pass not because they assert properties that are true, but because they assert nothing at all. We call such tests Rotten Green Tests. Rotten Green Tests represent a worst case: they report that the code under test is valid, but in fact do nothing to test that validity, beyond checking that the code does not crash. We describe an approach to identify rotten green tests by combining simple static and dynamic analyses. Our approach takes into account test helper methods, inherited helpers, and trait compositions, and has been implemented in a tool called DrTest. We have applied DrTest to several test suites in Pharo 7.0, and identified many rotten tests, including some that have been sleeping in Pharo for at least 5 years. [22]

Mining inline cache data to order inferred types in dynamic languages. The lack of static type information in dynamically-typed languages often poses obstacles for developers. Type inference algorithms can help, but inferring precise type information requires complex algorithms that are often slow. A simple approach that considers only the locally used interface of variables can identify potential classes for variables, but popular interfaces can generate a large number of false positives. We propose an approach called inline-cache type inference (ICTI) to augment the precision of fast and simple type inference algorithms. ICTI uses type information available in the inline caches during multiple software runs, to provide a ranked list of possible classes that most likely represent a variable's type. We evaluate ICTI through a proof-of-concept that we implement in Pharo Smalltalk. The analysis of the top-n+2 inferred types (where n is the number of recorded run-time types for a variable) for 5486 variables from four different software systems shows that ICTI produces promising results for about 75% of the variables. For more than 90% of variables, the correct run-time type is present among first six inferred types. Our ordering shows a twofold improvement when compared with the unordered basic approach, i.e., for a significant number of variables for which the basic approach offered ambiguous results, ICTI was able to promote the correct type to the top of the list. [22]

7.5. Blockchain

Ethereum Query Language Blockchains store a massive amount of heterogeneous data which will only grow in time. When searching for data on the Ethereum platform, one is required to either access the records (blocks) directly by using a unique identifier, or sequentially search several records to find the desired information. Therefore, we propose the Ethereum Query Language (EQL), a query language that allows users to retrieve information from the blockchain by writing SQL-like queries. The queries provide a rich syntax to specify data elements to search information scattered through several records. We claim that EQL makes it easier to search, acquire, format, and present information from the blockchain. [7]

SmartInspect: solidity smart contract inspector. Solidity is a language used for smart contracts on the Ethereum blockchain. Smart contracts are embedded procedures stored with the data they act upon. Debugging smart contracts is a really difficult task since once deployed, the code cannot be reexecuted and inspecting a simple attribute is not easily possible because data is encoded. We address the lack of inspectability of a deployed contract by analyzing contract state using decompilation techniques driven by the contract structure definition. Our solution, SmartInspect, also uses a mirror-based architecture to represent locally object responsible for the interpretation of the contract state. SmartInspect allows contract developers to better visualize and understand the contract stored state without needing to redeploy, nor develop any ad-hoc code. [8]

Preliminary Steps Towards Modeling Blockchain Oriented Software Even though blockchain is mostly popular for its cryptocurrency, smart contracts have become a very prominent blockchain application. Smart contracts are like classes that can be called by client applications outside the blockchain. Therefore it is possible to develop blockchain-oriented software (BOS) that implements part of the business logic in the blockchain by using smart contracts. Currently, there is no design standard to model BOS. Since modeling is an important part of designing a software, developers may struggle to plan their BOS. We show three complementary modeling approaches based on well-known software engineering models and apply them to a BOS example. Our goal is to start the discussion on specialized blockchain modeling notations. [13]

SmartAnvil: Open-Source Tool Suite for Smart Contract Analysis. Smart contracts are new computational units with special properties: they act as classes with aspectual concerns; their memory structure is more complex than mere objects; they are obscure in the sense that once deployed it is difficult to access their internal state; they reside in an append-only chain. There is a need to support the building of new generation tools to help developers. Such support should tackle several important aspects: (1) the static structure of the contract, (2) the object nature of published contracts, and (3) the overall data chain composed of blocks and transactions. In this chapter, we present SmartAnvil an open platform to build software analysis tools around smart contracts. We illustrate the general components and we focus on three important aspects: support for static analysis of Solidity smart contracts, deployed smart contract binary analysis through inspection, and blockchain navigation and querying. SmartAnvil is open-source and supports a bridge to the Moose data and software analysis platform. [21]

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. BlockChain

Participants: Henrique Rocha, Marcus Denker, Stéphane Ducasse
From 2016, ongoing.

We started a new collaboration with a local startup (UTOCAT) 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 started in 2017.

8.1.2. Pharo Consortium

Participants: Esteban Lorenzano, Clément Béra, Marcus Denker, Stéphane Ducasse
From 2012, ongoing.

The Pharo Consortium was founded in 2012 and is growing constantly. By the end 2018, it has 32 company members, 17 academic partners. Inria supports the consortium with one full time engineer starting in 2011. In 2018, the Pharo Consortium joined InriaSoft.

More at http://consortium.pharo.org.

8.2. Bilateral Grants with Industry

8.2.1. Thales CIFRE

Participants: Brice Govin, Anne Etien, Nicolas Anquetil, Stéphane Ducasse
From 2015, ongoing.

We are working on large industrial project rearchitecturization. PhD in progress: Brice Govin, *Support to implement a rejuvenated software architecture in legacy software*. CIFRE Thales started Jan 2015.

8.2.2. Remodularization of Architecture

Participants: Nicolas Anquetil, Santiago Bragagnolo Stéphane Ducasse, Anne Etien, Benoît Verhaeghe
From 2017, ongoing.

We started a new collaboration with the software editor Berger Levrault about software architecture remodularization. The collaboration started with an end study project exploring the architecture used in the company in order to later migrate from GWT to Angular JS since GWT will not be backward supported anymore in the next versions. An internship and a PhD CIFRE thesis will start in 2018.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. CAR IMT Douai

Participants: Pablo Tesone, Guillermo Polito, Marcus Denker, Stéphane Ducasse with: L. Fabresse and N. Bouraqadi (IMT Douai)
From 2009, ongoing.

We have signed a convention with the CAR team led by Noury Bouraqadi of IMT Douai. In this context we co-supervised three PhD students (Mariano Martinez-Peck, Nick Papoylias and Guillermo Polito). The team is also an important contributor and supporting organization of the Pharo project.

Currently, Pablo Tesone is doing a PhD co-supervided by RMOD and Pr. L. Fabresse and N. Bouraqadi. We are collaborating in the Context of CPER Data since 2018.

9.1.2. CPER DATA

Participants: Marcus Denker, Stéphane Ducasse, Allex Oliveira with: L. Fabresse and N. Bouraqadi (IMT Douai)
From 2018, ongoing.

Funding to work one year on the PharoThings Platform. We are creating content for a website and a Demo in collaboration with IMT Douai.

9.2. National Initiatives

9.2.1. CEA List

Participants: Jason Lecerf, Stéphane Ducasse with T. Goubier (CEA List)
From 2016, ongoing.

Jason Lecerf started a shared PhD Oct 2016: *Reuse of code artifacts for embedded systems through refactoring*.

9.3. European Initiatives

9.3.1. Collaborations in European Programs, Except FP7 & H2020

Namur University, Belgium
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.

VUB Brussels, Belgium
Participants: Guillermo Polito, Stéphane Ducasse.

From 2016, ongoing.

Student: Matteo Marra, collaboration with Eliza Gonzalez Boix. Guillermo Polito co-supervised Matteo Marra's master thesis.

University of Prague
Participants: Stéphane Ducasse.

From 2015, ongoing.

We are working with Dr. Robert Pergl from the University of Prague. Stéphane Ducasse gave a lecture at the University of Prague in 2018.

9.4. International Initiatives

9.4.1. Inria International Partners

9.4.1.1. Informal International Partners

Uqbar Argentina

Participants: Pablo Tesone, Esteban Lorenzano, Guillermo Polito, Stéphane Ducasse.

From 2015, ongoing.

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).

Pharo in Research:

Participants: Pablo Tesone, Esteban Lorenzano, Guillermo Polito, Marcus Denker, Stéphane Ducasse.
From 2009, ongoing.

We are building an ecosystem around Pharo with international research groups, universities and companies. Several research groups (such as Software Composition Group – Bern, and Pleaid – Santiago) are using Pharo. Many universities are teaching OOP using Pharo and its books. Several companies worldwide are deploying business solutions using Pharo.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

- Abdelghani Alidra [University of Skikda Algeria, from Nov 2018]
- Andy Amoordon [Utocat, from Oct 2018]
- Jan Bliznicenko [University of Prague , Mar 2018]
- Abdelhakim Bouremel [University of Skikda Algeria, May 2018]
- Thomas Dupriez [Ecole Normale Supŕieure Paris, Sep 2018]
- Christopher Fuhrman [Ecole de technologie superieure de montreal, from Oct 2018]
- Tomohiro Oda [Key Technology Laboratory, Japan, from Aug 2018 until Sep 2018]
- Giuseppe Antonio Pierro [University of Cagliari, from Aug 2018]
- Ronie Salgado Faila [niversity of Chile at Santiago, Chile, from Aug 2018 until Sep 2018]

- Serge Demeyer [Universiteit Antwerpen, Belgium, from Jun 2018 until Jul 2018

9.5.1.1. Internships

- Lionel Akue [Inria, from Jul 2018 until Sep 2018]
- Asbathou Biyalou Sama [Inria, from Apr 2018 until Aug 2018]
- Quentin Ducasse [Inria, from Jun 2018 until Aug 2018]
- Thomas Dupriez [Ecole Normale Supŕieure Cachan, until Feb 2018]
- Yoan Geran [Ecole Normale Supŕieure Paris, from Jun 2018 until Jul 2018]
- Pierre Tsapliayeu [Univ de Lille, from Apr 2018 until Aug 2018]
- Eleonore Wald [Univ de Lille, from Apr 2018 until Jul 2018]
- Oleksandr Zaitsev [Inria, from Oct 2018]
- Myroslava Romaniuk [Inria, from Jul 2018 until Sep 2018]

9.5.2. Visits to International Teams

- Marcus Denker: VUB Brussels in spring and fall 2018 (Lecture).

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. Member of the Organizing Committees

- Marcus Denker and Stéphane Ducasse are in the board of ESUG and organized ESUG 2018, the yearly Smalltalk conference that brings together research and industry (http://www.esug.org/).

10.1.2. Scientific Events Selection

10.1.2.1. Chair of Conference Program Committees

Anne Etien has been PC chair of IWST since 2015.

10.1.2.2. Reviewer

Nicolas Anquetil: SCAM'18, the 18th IEEE International Working Conference on Source Code Analysis and Manipulation.

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

Nicolas Anquetil reviewed articles for the following international journals:

- IEEE Transactions on Software Engineering;
- Information and Software Technology;
- Journal of Systems and Software.

10.1.4. Scientific Expertise

- Anne Etien: expert for the French government on Research tax credit.
- Marcus Denker: reviewer for EUREKA.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master : Christophe Demarey, Intégration continue, 16 EdTD, M2, Université de Lille, France

Licence: Anne Etien, Bases de données, 30h, L3, Polytech-Lille, France

Licence: Anne Etien, Programmation par objet, 40h, L3, Polytech-Lille, France

Master: Anne Etien, Metamodelisation for Reverse Engineering, 5h, M2, Université de Montpellier, France

Master: Anne Etien, Metamodelisation for Reverse Engineering, 12h, M2, Université Paris 1, France

Master: Anne Etien, Metamodelisation for Reverse Engineering, 5h, M2, Tunis, Tunisie

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, Syste`me d'information objet, 10h, M1, Polytech Lille, France

Master: Anne Etien, Bases de donnés Avancés, 20h, M1, Polytech-Lille, France

Master: Anne Etien, Qualité logicielle, 8h, M2, Université Lille 1, France

Licence : Vincent Aranega, Initiation à l'informatique, 51 EdTD, L1, Université de Lille, France

Licence : Vincent Aranega, Architecture des ordinateurs, 42 EdTD, L2, Université de Lille, France

Licence : Vincent Aranega, Génie Logiciel, 42 EdTD, M1, Université de Lille, France

License : Damien Pollet, OpenDevs, 35h, L3, Université de Lille, France

License : Damien Pollet, Programmation objet en Java, 155h, L3, IMT Lille-Douai, France

License : Damien Pollet, Systèmes numériques, 14h, L3, IMT Lille-Douai, France

Master : Damien Pollet, Technologies des systèmes d'informations, 21h, M1, IMT Lille-Douai, France

Master : Damien Pollet, Ingénierie du logiciel, 7h, M2, IMT Lille-Douai, France

Master : Damien Pollet, Algorithmes pour les réseaux, 27h, M2, IMT Lille-Douai, France

Master: Meta modeling, Ecole des Mines de Douai, 22 EdTD, M1, Nov 2018, Douai, France

Master: Advanced Design, ENIS M1, Dec 2017/2018, Tunis, 22 EdTD, Tunisia

License: Object-oriented programming, Prague L3, Dec 2017/2018, 16 EdTD, Prague, Tcheque Republic

Master: Meta modeling, Pantheon Sorbonne, 27 EdTD, M2, Dec 2018, Paris, France

Master: Meta programming, Universite de Bretagne Occidental, 9 EdTD, brest, France.

Licence : Thomas Dupriez, Mathématiques Discrètes, 39 heures en équivalent TD, L2, Université Lille, France

Master : Marcus Denker, 2 hours (spring and fall each), Advanced Reflection. MetaLinks, VUB Brussels, Belgium.

License: open-dev L3 option 17 hours Lille, France

License: Introduction à la Programmation L3 option 31.5 hours Paul Sabatier Toulouse, France

E-learning

Pharo Mooc, 7 weeks, Licences and Master students

Pedagogical resources: Books, Learning Object-Oriented Programming, Design and TDD with Pharo, A simple reflective object kernel.

10.2.2. Supervision

PhD: Brice Govin, *Support to Implement a Rejuvenated Software Architecture in Legacy Software* CIFRE Thales, 26 June 2018, Anne Etien, Nicolas Anquetil. [1]

PhD: Pablo Tesone, *Hot Software Update In Robotics Applications*, IMT Lille-Douai, 16 December, Luc Fabresse, Stéphane Ducasse

PhD in progress: Jason Lecerf, *Reuse of Code Artifacts for Embedded Systems Through Refactoring*, started Oct 2016, CEA Thierry Goubier, Stéphane Ducasse

PhD in progress: Julien Delplanque, *Software Engineering Techniques Applied to Databases*, started Oct 2017, Anne Etien, Nicolas Anquetil

PhD in progress: Lionel Akue, *Fortran Analysis*, started Oct 2018, Anne Etien, Nicolas Anquetil

PhD in progress: Thomas Dupriez, *New Generation Debugger and Application Monitoring*, started Oct 2018, Stephane Ducasse, Guillermo Polito

PhD in progress: Houekpetodji Mahugnon Honore, *Multi-Facet Actionable for Information System Rejuvenation*, SPI Lille, France, Stephane Ducasse, Nicolas Anquetil, Nicolas Dias, Jerome Sudich

PhD in progress: Carolina Hernandez, *Tools for MicroKernels* Guillermo Polito and Luc Fabresse

10.3. Popularization

10.3.1. Articles and contents

- Stéphane Ducasse and Guillermo Polito. Stéphane Ducasse (Ed.). Physche: A Little Scheme in Pharo, p. 50, Square Bracket Associates, 2018. [20]
- Stéphane Ducasse. Stéphane Ducasse (Ed.). A Simple Reflective Object Kernel, p. 40, Square Bracket Associates, 2018. [19]

10.3.2. Education

- A MOOC for Pharo is online (Stéphane Ducasse).
  http://mooc.pharo.org

10.3.3. Interventions

- Multiple public Pharo Sprints in Lille.
  https://association.pharo.org/events
- IoT Hackathon. One Day Event in Cologne, Germany. Using PharoIoT
  http://zweidenker.de/de/iot-hackathon-2018
  Video: https://www.youtube.com/watch?v=dIl9FAatKyw
- RMOD co-organized and participated at ESUG 2018.
  http://www.esug.org/wiki/pier/Conferences/2018
  - Presentation about PharoIoT.
    https://www.slideshare.net/MarcusDenker/pharo-iot

11. Bibliography

Publications of the year

Doctoral Dissertations and Habilitation Theses

[1] B. Govin.*Supporting Legacy Software Architecture Renovation : A real Case by Thales Air Systems*, Lille, June 2018, https://hal.inria.fr/tel-01881319

Articles in International Peer-Reviewed Journal

[2] G. Chari, D. Garbervetsky, S. Marr, S. Ducasse.*Fully Reflective Execution Environments: Virtual Machines for More Flexible Software*, in "IEEE Transactions on Software Engineering", May 2018, p. 1 - 20 [*DOI :* 10.1109/TSE.2018.2812715], https://hal.inria.fr/hal-01728111

[3] A. Etien, J. Laval.*Advances in Dynamic Languages*, in "Science of Computer Programming", September 2018, vol. 161, 1, cited By 0 [*DOI :* 10.1016/j.scico.2018.03.004], https://hal.archives-ouvertes.fr/hal-01858483

[4] M. Marra, G. Polito, E. Gonzalez Boix.*Out-Of-Place debugging: a debugging architecture to reduce debugging interference*, in "The Art, Science, and Engineering of Programming", November 2018, vol. 3, n° 2, p. 1-29 [*DOI :* 2019/3/3], https://hal.inria.fr/hal-01952790

[5] N. Milojković, C. Béra, M. Ghafari, O. Nierstrasz.*Mining inline cache data to order inferred types in dynamic languages*, in "Science of Computer Programming", September 2018, vol. 161, p. 105-121 [*DOI :* 10.1016/j.scico.2017.11.003], https://hal.inria.fr/hal-01666541

[6] P. Tesone, G. Polito, N. Bouraqadi, S. Ducasse, L. Fabresse.*Dynamic Software Update from Development to Production*, in "The Journal of Object Technology", November 2018, vol. 17, n° 1, p. 1-36 [*DOI :* 10.5381/jot.2018.17.1.a2], https://hal.inria.fr/hal-01920362

International Conferences with Proceedings

[7] S. Bragagnolo, H. S. C. Rocha, M. Denker, S. Ducasse.*Ethereum Query Language*, in "WETSEB 2018 - 1st International Workshop on Emerging Trends in Software Engineering for Blockchain", Gothenburg, Sweden, May 2018 [*DOI :* 10.1145/3194113.3194114], https://hal.inria.fr/hal-01831084

[8] S. Bragagnolo, H. S. C. Rocha, M. Denker, S. Ducasse.*SmartInspect: Solidity Smart Contract Inspector*, in "IWBOSE 2018 - 1st International Workshop on Blockchain Oriented Software Engineering", Campobasso, Italy, IEEE, March 2018 [*DOI :* 10.1109/IWBOSE.2018.8327566], https://hal.inria.fr/hal-01831075

[9] S. Costiou, M. Kerboeuf, A. Plantec, M. Denker.*Collectors*, in "Programming Experience 2018 (PX'18)", Nice, France, Companion of the 2nd International Conference on Art, Science, and Engineering of Programming, ACM Press, April 2018, 9 [*DOI :* 10.1145/3191697.3214335], https://hal.univ-brest.fr/hal-01829183

[10] J. Delplanque, A. Etien, N. Anquetil, O. Auverlot.*Relational Database Schema Evolution: An Industrial Case Study*, in "ICSME 2018 - 34th IEEE International Conference on Software Maintenance and Evolution", Madrid, Spain, September 2018 [*DOI :* 10.1109/ICSME.2018.00073], https://hal.archives-ouvertes.fr/hal-01945042

[11] S. Demeyer, B. Verhaeghe, A. Etien, N. Anquetil, S. Ducasse.*Evaluating the Efficiency of Continuous Testing during Test-Driven Development*, in "VST 2018 - 2nd IEEE International Workshop on Validation, Analysis and Evolution of Software Tests", Campobasso, Italy, March 2018, p. 1-5 [*DOI :* 10.1109/VST.2018.8327152], https://hal.inria.fr/hal-01717343

[12] J. LECERF, J. BRANT, T. GOUBIER, S. DUCASSE.*A Reflexive and Automated Approach to Syntactic Pattern Matching in Code Transformations*, in "ICSME 2018 - 34th IEEE International Conference on Software Maintenance and Evolution", Madrid, Spain, ICSME 2018 - 34th IEEE International Conference on Software Maintenance and Evolution, September 2018 [*DOI :* 10.1109/ICSME.2018.00052], https://hal.archives-ouvertes.fr/hal-01851857

[13] H. ROCHA, S. DUCASSE.*Preliminary Steps Towards Modeling Blockchain Oriented Software*, in "WETSEB 2018 - 1st International Workshop on Emerging Trends in Software Engineering for Blockchain", Gothenburg, Sweden, May 2018 [*DOI :* 10.1145/3194113.3194123], https://hal.inria.fr/hal-01831046

[14] C. M. SOUZA COUTO, H. ROCHA, R. TERRA.*A Quality-oriented Approach to Recommend Move Method Refactorings*, in "Proceedings of the 17th Brazilian Symposium on Software Quality", Curitiba, Brazil, October 2018 [*DOI :* 10.1145/3275245.3275247], https://hal.inria.fr/hal-01944493

[15] P. TESONE, G. POLITO, L. FABRESSE, N. BOURAQADI, S. DUCASSE.*Implementing Modular Class-based Reuse Mechanisms on Top of a Single Inheritance VM*, in "SAC 2018:- The 33rd ACM/SIGAPP Symposium On Applied Computing", Pau, France, SAC 2018 Symposium on Applied Computing, ACM New York, NY, USA, April 2018 [*DOI :* 10.1145/3167132.3167244], https://hal.archives-ouvertes.fr/hal-01812612

[16] M. VIGGIATO, R. TERRA, H. S. C. ROCHA, M. TULIO VALENTE, E. FIGUEIREDO.*Microservices in Practice: A Survey Study*, in "VEM 2018 - 6th Workshop on Software Visualization, Evolution and Maintenance", Sao Carlos, Brazil, September 2018, https://hal.inria.fr/hal-01944464

Conferences without Proceedings

[17] J. DELPLANQUE, O. AUVERLOT, A. ETIEN, N. ANQUETIL.*Définition et identification des tables de nomenclatures*, in "INFORSID 2018 - 36ème édition d'INFormatique des ORganisations et Systèmes d'Information et de Décision", Nantes, France, May 2018, https://hal.archives-ouvertes.fr/hal-01944135

[18] S. KALEBA, C. BÉRA, S. DUCASSE.*Assessing primitives performance on multi-stage execution*, in "ICOOOLPS 2017 - 12th Workshop on Implementation, Compilation, Optimization of Object-Oriented Languages, Programs and Systems", Amsterdam, Netherlands, July 2018, https://hal.archives-ouvertes.fr/hal-01874946

Scientific Books (or Scientific Book chapters)

[19] S. DUCASSE.*A simple reflective object kernel*, Published by the author, September 2018, https://hal.inria.fr/hal-01900323

[20] S. DUCASSE, G. POLITO.*Physche: A Little Scheme in Pharo*, Published by the authors, August 2018, https://hal.inria.fr/hal-01900327

[21] S. DUCASSE, H. ROCHA, S. BRAGAGNOLO, M. DENKER, C. FRANCOMME.*SmartAnvil: Open-Source Tool Suite for Smart Contract Analysis*, in "Blockchain and Web 3.0: Social, economic, and technological challenges", Routledge, February 2019, https://hal.inria.fr/hal-01940287

Research Reports

[22] J. DELPLANQUE, S. DUCASSE, A. BLACK, G. POLITO.*Rotten Green Tests A First Analysis*, Inria Lille Nord Europe - Laboratoire CRIStAL - Université de Lille ; Portland State University, Oregon, USA, June 2018, https://hal.archives-ouvertes.fr/hal-01819302

[23] M. DENKER, N. ANQUETIL, S. DUCASSE, A. ETIEN, D. POLLET.*Project-Team RMoD (Analyses and Language Constructs for Object-Oriented Application Evolution) 2017 Activity Report*, Inria Lille - Nord Europe, January 2018, https://hal.inria.fr/hal-01683649

References in notes

[24] 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

[25] 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

[26] 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

[27] 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

[28] 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

[29] 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

[30] 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

[31] 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

[32] P. COINTE.*Metaclasses are First Class: the ObjVlisp Model*, in "Proceedings OOPSLA '87, ACM SIGPLAN Notices", December 1987, vol. 22, p. 156–167

[33] 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

[34] 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

[35] 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

[36] 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° 2, p. 331–388 [*DOI :* 10.1145/1119479.1119483], http://scg.unibe.ch/archive/papers/Duca06bTOPLASTraits.pdf

[37] 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

[38] 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

[39] K. FISHER, J. REPPY.*Statically typed traits*, University of Chicago, Department of Computer Science, December 2003, n° TR-2003-13

[40] P. W. L. FONG, C. ZHANG.*Capabilities as alias control: Secure cooperation in dynamically extensible systems*, Department of Computer Science, University of Regina, 2004

[41] M. FURR, J.-H. AN, J. S. FOSTER.*Profile-guided static typing for dynamic scripting languages*, in "OOPSLA'09", 2009

[42] A. GOLDBERG.*Smalltalk 80: the Interactive Programming Environment*, Addison Wesley, Reading, Mass., 1984

[43] L. GONG.*New security architectural directions for Java*, in "compcon", 1997, vol. 0, 97, http://dx.doi.org/10.1109/CMPCON.1997.584679

[44] M. HICKS, S. NETTLES.*Dynamic software updating*, in "ACM Transactions on Programming Languages and Systems", nov 2005, vol. 27, n° 6, p. 1049–1096, http://dx.doi.org/10.1145/1108970.1108971

[45] G. KICZALES, J. DES RIVIÈRES, D. G. BOBROW.*The Art of the Metaobject Protocol*, MIT Press, 1991

[46] G. KICZALES, L. RODRIGUEZ.*Efficient Method Dispatch in PCL*, in "Proceedings of ACM conference on Lisp and Functional Programming", Nice, 1990, p. 99–105

[47] 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

[48] S. LIANG, G. BRACHA. *Dynamic Class Loading in the Java Virtual Machine*, in "Proceedings of OOPSLA '98, ACM SIGPLAN Notices", 1998, p. 36–44

[49] 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

[50] 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

[51] R. C. MARTIN. *Agile Software Development. Principles, Patterns, and Practices*, Prentice-Hall, 2002

[52] M. S. MILLER. *Robust Composition: Towards a Unified Approach to Access Control and Concurrency Control*, Johns Hopkins University, Baltimore, Maryland, USA, May 2006

[53] 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

[54] 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

[55] P. J. QUITSLUND. *Java Traits — Improving Opportunities for Reuse*, OGI School of Science & Engineering, Beaverton, Oregon, USA, September 2004, n^o CSE-04-005

[56] J. REPPY, A. TURON. *A Foundation for Trait-based Metaprogramming*, in "International Workshop on Foundations and Developments of Object-Oriented Languages", 2006

[57] F. RIVARD. *Pour un lien d'instanciation dynamique dans les langages à classes*, in "JFLA96", Inria — collection didactique, January 1996

[58] 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

[59] 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

[60] 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

[61] 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

[62] C. SMITH, S. DROSSOPOULOU.*Chai: Typed Traits in Java*, in "Proceedings ECOOP 2005", 2005

[63] G. SNELTING, F. TIP.*Reengineering Class Hierarchies using Concept Analysis*, in "ACM Trans. Programming Languages and Systems", 1998

[64] K. J. SULLIVAN, W. G. GRISWOLD, Y. CAI, B. HALLEN.*The Structure and Value of Modularity in SOftware Design*, in "ESEC/FSE 2001", 2001

[65] D. VAINSENCHER.*MudPie: layers in the ball of mud*, in "Computer Languages, Systems & Structures", 2004, vol. 30, n^{o} 1-2, p. 5–19

[66] 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

1. Team, Visitors, External Collaborators **363**
2. Overall Objectives **364**
3. Research Program **365**
3.1. In Short 365
3.2. Decision-making Under Uncertainty 365
3.2.1. Reinforcement Learning 366
3.2.2. Multi-arm Bandit Theory 367
3.3. Statistical analysis of time series 368
3.3.1. Prediction of Sequences of Structured and Unstructured Data 368
3.3.2. Hypothesis testing 368
3.3.3. Change Point Analysis 369
3.3.4. Clustering Time Series, Online and Offline 369
3.3.5. Online Semi-Supervised Learning 369
3.3.6. Online Kernel and Graph-Based Methods 369
4. Application Domains **370**
5. Highlights of the Year **370**
6. New Software and Platforms **370**
6.1. BAC 370
6.2. GuessWhat?! 371
6.3. Squeak 371
6.4. OOR 371
6.5. DPPy 371
6.6. SMPyBandits 372
7. New Results **372**
7.1. Decision-making Under Uncertainty 372
7.1.1. Reinforcement Learning 372
7.1.2. Multi-armed Bandit Theory 374
7.1.3. Stochastic Games 377
7.1.4. Online Kernel and Graph-Based Methods 377
7.2. Applications 378
7.2.1. Dialogue Systems and Natural Language 378
7.2.2. Recommendation systems 379
7.2.3. Autonomous car 379
7.2.4. Software development 379
7.2.5. Deep Learning 380
8. Bilateral Contracts and Grants with Industry **381**
8.1.1. Lelivrescolaire.fr 381
8.1.2. Sidexa 381
8.1.3. Renault 381
8.1.4. Critéo 381
8.1.5. Orange Labs 382
8.1.6. 55 382
8.1.7. AB-Tasty 382
9. Partnerships and Cooperations **382**
9.1. National Initiatives 382
9.1.1. ANR BoB 382
9.1.2. ANR Badass 383
9.1.3. ANR ExTra-Learn 384
9.1.4. Grant of Fondation Mathématique Jacques Hadamard 385

9.1.5. National Partners 385
9.2. European Initiatives 385
9.2.1.1. H2020 BabyRobot 385
9.2.1.2. CHIST-ERA DELTA 386
9.2.1.3. CHIST-ERA IGLU 387
9.3. International Initiatives 387
9.3.1. Inria Associate Teams Not Involved in an Inria International Labs 387
9.3.2. Inria International Partners 388
9.3.2.1. Declared Inria International Partners 388
9.3.2.2. CWI 388
9.3.3. Participation in Other International Programs 388
9.4. International Research Visitors 389
9.4.1. Visits of International Scientists 389
9.4.2. Visits to International Teams 390
10. Dissemination **390**
10.1. Promoting Scientific Activities 390
10.1.1. Scientific Events Organisation 390
10.1.1.1. EWRL 2018 390
10.1.1.2. General Chair, Scientific Chair 390
10.1.1.3. Member of the Organizing Committees 390
10.1.2. Scientific Events Selection 390
10.1.2.1. Member of the Conference Program Committees 390
10.1.2.2. Reviewer 390
10.1.3. Journal 391
10.1.4. Invited Talks 391
10.1.5. Scientific Expertise 392
10.1.6. Research Administration 392
10.2. Teaching - Supervision - Juries 392
10.2.1. Teaching 392
10.2.2. Supervision 393
10.2.3. Juries 393
10.3. Popularization 394
10.3.1. Internal or external Inria responsibilities 394
10.3.2. Articles and contents 394
10.3.3. Education 394
10.3.4. Interventions 394
10.3.5. Creation of media or tools for science outreach 394
11. Bibliography **394**

Project-Team SEQUEL

Creation of the Project-Team: 2007 July 01

Keywords:

Computer Science and Digital Science:

A3. - Data and knowledge
A3.1. - Data
A3.1.1. - Modeling, representation
A3.1.4. - Uncertain data
A3.3. - Data and knowledge analysis
A3.3.1. - On-line analytical processing
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.3. - Reinforcement learning
A3.4.4. - Optimization and learning
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A3.5.2. - Recommendation systems
A5.1. - Human-Computer Interaction
A9. - Artificial intelligence
A9.2. - Machine learning
A9.3. - Signal analysis
A9.4. - Natural language processing
A9.7. - AI algorithmics

Other Research Topics and Application Domains:

B5.8. - Learning and training
B6.1. - Software industry
B7.2.1. - Smart vehicles
B9.1.1. - E-learning, MOOC
B9.5. - Sciences
B9.5.6. - Data science

1. Team, Visitors, External Collaborators

Research Scientists

Émilie Kaufmann [CNRS, Researcher]
Odalric Maillard [Inria, Researcher]
Michal Valko [Inria, Researcher, HDR]

Faculty Member

Philippe Preux [Team leader, Univ Charles de Gaulle, Professor, HDR]

External Collaborators

Remi Bardenet [CNRS]
Jérémie Mary [Critéo, HDR]
Olivier Pietquin [Google, HDR]

Technical Staff

Guillaume Gautier [CNRS&Inria, since 2017]

PhD Students

Sheikh Waqas Akhtar [Inria, until Aug 2018]
Merwan Barlier [Univ de Lille, from Feb 2018 until Nov 2018]
Lilian Besson [Ecole Normale Supérieure Cachan]
Nicolas Carrara [Orange Labs puis Univ de Lille, since 2015]
Omar Darwiche Domingues [Inria, from Oct 2018]
Yannis Flet Berliac [Univ des sciences et technologies de Lille, from Oct 2018]
Ronan Fruit [Inria, since Dec 2015]
Jean-Bastien Grill [Ecole Normale Supérieure Paris, until Dec 2018]
Édouard Leurent [Renault]
Pierre Perrault [Inria]
Hassan Saber [Inria, from Oct 2018]
Mathieu Seurin [Univ des sciences et technologies de Lille]
Julien Seznec [Le Livre Scolaire]
Xuedong Shang [Univ des sciences et technologies de Lille]
Florian Strub [Univ des sciences et technologies de Lille]
Kiewan Villatel [Critéo]
Romain Warlop [55, until Oct 2018]

Post-Doctoral Fellows

Matteo Pirotta [Inria, until Oct 2018]
Mohammad Sadegh Talebi Mazraeh Shahi [Inria, from Jun 2018]

Visiting Scientist

Jumpei Komiyama [Univ Tokyo, from Sep 2018 until Nov 2018]

Administrative Assistant

Amelie Supervielle [Inria]

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 [62]). 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 [60].

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:

$$V^\pi(x) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t | x_0 = x, \pi\right], \tag{5}$$

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 [59]) 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.

In order to maximize a given functional in a sequential framework, one usually applies Dynamic Programming (DP) [57], 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 [58]. 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 $x \in \mathcal{X}$,

$$\pi(x) \in \arg\max_{a \in \mathcal{A}} \sum_{x' \in \mathcal{X}} p(x'|x, a)\left[r(x, a, x') + \gamma V(x')\right].$$

[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.

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 $x \in \mathcal{X}$,

$$V^*(x) = \max_{a \in \mathcal{A}} \sum_{x' \in \mathcal{X}} p(x'|x, a) \left[r(x, a, x') + \gamma V^*(x') \right]. \tag{6}$$

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:

$$\pi^*(x) \in \arg\max_{a \in \mathcal{A}} \sum_{x' \in \mathcal{X}} p(x'|x, a) \left[r(x, a, x') + \gamma V^*(x') \right]. \tag{7}$$

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 ([63]):

- 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 [61], 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.* [56] 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 behavior (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 behavior, or than a class of other behaviors.

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) behavior 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 $x^1 = (x^1_1, \cdots, x^1_{n_1}), \cdots, x^N = (x^1_N, \cdots, x^N_{n_N})$, 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 advertising and recommendation 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 recommendation 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

- Daniele Calandriello is awarded with the AFIA price for his PhD defended in December 2017. As a side note, this is the 5th time a PhD student of SequeL receives this award since our first PhD defense in 2010.
- We organized the 14th European Workshop on Reinforcement Learning in Lille. This event gathered 200 researchers; there were a dozen invited presentations by world research leaders, including Prof. Richard Sutton (U. Alberta), the founder of modern RL, Prof. Tze Leung Lai (Stanford U.), one of the key reference in bandit research, and also Nicolò Cesa-Bianchi (U. Milan), Peter Auer (U. of Leoben), Rémi Munos (Deepmind Paris), and Joelle Pineau (Mc Gill and FAIR).

5.1.1. Awards

- Former 2017 intern M. Asadi got a It was “Best Poster Award” at Transylvania Machine Learning Summer School (TMLSS), July 2018 for the work she did while in SequeL
- É. Kaufmann is among the top 10 reviewers at ICML 2018 (out of 1800 reviewers)
- Ph. Preux was among the 24 “level-2 Distinguished Senior Program Committee Members” for IJCAI 2018 (out of 498 SPC members, 115 were distinguished, 23 at level 2, the highest level)
- M. Valko is among the top 10 reviewers at ICML 2018 (out of 1800 reviewers)

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. GuessWhat?!

GuessWhat?! Visual object discovery through multi-modal dialogue
KEYWORDS: Deep learning - Dialogue System
FUNCTIONAL DESCRIPTION: This project train a AI to play the GuessWhat?! game. Thus, you can train an AI to ask questions, to answer questions about images. You can also perform basic visual reasoning. This project is a testbed for future interactive dialogue system.

- Partner: Universite de Montreal
- Contact: Florian Strub
- Publications: GuessWhat?! Visual object discovery through multi-modal dialogue - End-to-end optimization of goal-driven and visually grounded dialogue systems Harm de Vries

6.3. Squeak

Sequential sampling for kernel matrix approximation
KEYWORD: Machine learning

- Contact: Daniele Calandriello
- URL: http://researchers.lille.inria.fr/~valko/hp/serve.php?what=publications/squeak.py

6.4. OOR

Optimistic Optimization in R
KEYWORDS: Black-box optimization - Machine learning

- Contact: Mickael Binois
- URL: https://cran.r-project.org/web/packages/OOR/index.html

6.5. DPPy

Sampling Determinantal Point Processes with Python
KEYWORD: Determinantal point processes
FUNCTIONAL DESCRIPTION: Determinantal point processes (DPPs) are specific probability distributions over clouds of points that are used as models and computational tools across physics, probability, statistics, and more recently machine learning. Sampling from DPPs is nontrivial and therefore we present. DPPy, a Python toolbox that gathers known exact and approximate sampling algorithms. The project is hosted on GitHub and equipped with an extensive documentation.

- Contact: Guillaume Gautier
- URL: https://github.com/guilgautier/DPPy/

6.6. SMPyBandits

Open-Source Python package for Single- and Multi-Players multi-armed Bandits algorithms.
KEYWORD: Machine learning
FUNCTIONAL DESCRIPTION: The library contains the implementation of many single-player mutli-armed bandit algorithms as well as the implementation of all the state-of-the-art multi-player algorithms.

- Contact: Lilian Besson

7. New Results

7.1. Decision-making Under Uncertainty

7.1.1. Reinforcement Learning

A Fitted-Q Algorithm for Budgeted MDPs, [26]

We address the problem of budgeted reinforcement learning, in continuous state-space, using a batch of transitions. To this extend, we introduce a novel algorithm called Budgeted Fitted-Q (BFTQ). Benchmarks show that BFTQ performs as well as a regular Fitted-Q algorithm in a continuous 2-D world but also allows one to choose the right amount of budget that fits to a given task without the need of engineering the rewards. We believe that the general principles used to design BFTQ can be applied to extend others classical reinforcement learning algorithms for budgeted oriented applications.

Safe transfer learning for dialogue applications, [27]

In this paper, we formulate the hypothesis that the first dialogues with a new user should be handle in a very conservative way, for two reasons : avoid user dropout; gather more successful dialogues to speedup the learning of the asymptotic strategy. To this extend, we propose to transfer a safe strategy to initiate the first dialogues.

Variance-Aware Regret Bounds for Undiscounted Reinforcement Learning in MDPs, [17]

The problem of reinforcement learning in an unknown and discrete Markov Decision Process (MDP) under the average-reward criterion is considered, when the learner interacts with the system in a single stream of observations, starting from an initial state without any reset. We revisit the minimax lower bound for that problem by making appear the local variance of the bias function in place of the diameter of the MDP. Furthermore, we provide a novel analysis of the KL-Ucrl algorithm establishing a high-probability regret bound scaling as O S s,a V s,a T for this algorithm for ergodic MDPs, where S denotes the number of states and where V s,a is the variance of the bias function with respect to the next-state distribution following action a in state s. The resulting bound improves upon the best previously known regret bound O(DS $\sqrt{}$ AT) for that algorithm, where A and D respectively denote the maximum number of actions (per state) and the diameter of MDP. We finally compare the leading terms of the two bounds in some benchmark MDPs indicating that the derived bound can provide an order of magnitude improvement in some cases. Our analysis leverages novel variations of the transportation lemma combined with Kullback-Leibler concentration inequalities, that we believe to be of independent interest.

Efficient Bias-Span-Constrained Exploration-Exploitation in Reinforcement Learning, [29] We introduce SCAL, an algorithm designed to perform efficient exploration-exploitation in any unknown weakly-communicating Markov decision process (MDP) for which an upper bound c on the span of the optimal bias function is known. For an MDP with S states, A actions and $\Gamma \leq S$ possible next states, we prove a regret bound of $\widetilde{O}(c\sqrt{\Gamma SAT})$, which significantly improves over existing algorithms (e.g., UCRL and PSRL), whose regret scales linearly with the MDP diameter D. In fact, the optimal bias span is finite and often much smaller than D (e.g., $D = \infty$ in non-communicating MDPs). A similar result was originally derived by Bartlett and Tewari (2009) for REGAL.C, for which no tractable algorithm is available. In this paper, we relax the optimization problem at the core of REGAL.C, we carefully analyze its properties, and we provide the first computationally efficient algorithm to solve it. Finally, we report numerical simulations supporting our theoretical findings and showing how SCAL significantly outperforms UCRL in MDPs with large diameter and small span.

Near Optimal Exploration-Exploitation in Non-Communicating Markov Decision Processes, [28]

While designing the state space of an MDP, it is common to include states that are transient or not reachable by any policy (e.g., in mountain car, the product space of speed and position contains configurations that are not physically reachable). This leads to defining weakly-communicating or multi-chain MDPs. In this paper, we introduce TUCRL, the first algorithm able to perform efficient exploration-exploitation in any finite Markov Decision Process (MDP) without requiring any form of prior knowledge. In particular, for any MDP with $S^{\mathtt{C}}$ communicating states, A actions and $\Gamma^{\mathtt{C}} \leq S^{\mathtt{C}}$ possible communicating next states, we derive a $\widetilde{O}(D^{\mathtt{C}}\sqrt{\Gamma^{\mathtt{C}} S^{\mathtt{C}} AT})$ regret bound, where $D^{\mathtt{C}}$ is the diameter (i.e., the longest shortest path) of the communicating part of the MDP. This is in contrast with optimistic algorithms (e.g., UCRL, Optimistic PSRL) that suffer linear regret in weakly-communicating MDPs, as well as posterior sampling or regularized algorithms (e.g., REGAL), which require prior knowledge on the bias span of the optimal policy to bias the exploration to achieve sub-linear regret. We also prove that in weakly-communicating MDPs, no algorithm can ever achieve a logarithmic growth of the regret without first suffering a linear regret for a number of steps that is exponential in the parameters of the MDP. Finally, we report numerical simulations supporting our theoretical findings and showing how TUCRL overcomes the limitations of the state-of-the-art.

Upper Confidence Reinforcement Learning exploiting state-action equivalence, [53]

Stochastic Variance-Reduced Policy Gradient, [34]

In this paper, we propose a novel reinforcement-learning algorithm consisting in a stochastic variance-reduced version of policy gradient for solving Markov Decision Processes (MDPs). Stochastic variance-reduced gradient (SVRG) methods have proven to be very successful in supervised learning. However, their adaptation to policy gradient is not straightforward and needs to account for I) a non-concave objective function; II) approximations in the full gradient computation; and III) a non-stationary sampling process. The result is SVRPG, a stochastic variance-reduced policy gradient algorithm that leverages on importance weights to preserve the unbiasedness of the gradient estimate. Under standard assumptions on the MDP, we provide convergence guarantees for SVRPG with a convergence rate that is linear under increasing batch sizes. Finally, we suggest practical variants of SVRPG, and we empirically evaluate them on continuous MDPs.

Importance Weighted Transfer of Samples in Reinforcement Learning, [38]

We consider the transfer of experience samples (i.e., tuples < s, a, s', r >) in reinforcement learning (RL), collected from a set of source tasks to improve the learning process in a given target task. Most of the related approaches focus on selecting the most relevant source samples for solving the target task, but then all the transferred samples are used without considering anymore the discrepancies between the task models. In this paper, we propose a model-based technique that automatically estimates the relevance (importance weight) of each source sample for solving the target task. In the proposed approach, all the samples are transferred and used by a batch RL algorithm to solve the target task, but their contribution to the learning process is proportional to their importance weight. By extending the results for importance weighting provided in supervised learning literature, we develop a finite-sample analysis of the proposed batch RL algorithm. Furthermore, we empirically compare the proposed algorithm to state-of-the-art approaches, showing that it achieves better learning performance and is very robust to negative transfer, even when some source tasks are significantly different from the target task.

Training Dialogue Systems With Human Advice, [20]

One major drawback of Reinforcement Learning (RL) Spoken Dialogue Systems is that they inherit from the general exploration requirements of RL which makes them hard to deploy from an industry perspective. On the other hand, industrial systems rely on human expertise and hand written rules so as to avoid irrelevant behavior to happen and maintain acceptable experience from the user point of view. In this paper, we attempt to bridge the gap between those two worlds by providing an easy way to incorporate all kinds of human expertise in the training phase of a Reinforcement Learning Dialogue System. Our approach, based on the TAMER framework, enables safe and efficient policy learning by combining the traditional Reinforcement Learning reward signal with an additional reward, encoding expert advice. Experimental results show that our method leads to substantial improvements over more traditional Reinforcement Learning methods.

7.1.1.1. Deep reinforcement learning

FiLM: Visual Reasoning with a General Conditioning Layer, [35]

We introduce a general-purpose conditioning method for neural networks called FiLM: Feature-wise Linear Modulation. FiLM layers influence neural network computation via a simple , feature-wise affine transformation based on conditioning information. We show that FiLM layers are highly effective for visual reasoning - answering image-related questions which require a multi-step, high-level process - a task which has proven difficult for standard deep learning methods that do not explicitly model reasoning. Specifically, we show on visual reasoning tasks that FiLM layers 1) halve state-of-the-art error for the CLEVR benchmark, 2) modulate features in a coherent manner, 3) are robust to ablations and architectural modifications, and 4) generalize well to challenging, new data from few examples or even zero-shot.

Feature-wise transformations, [13]

Deep Reinforcement Learning and the Deadly Triad, [55]

We know from reinforcement learning theory that temporal difference learning can fail in certain cases. Sutton and Barto (2018) identify a deadly triad of function approximation, bootstrapping, and off-policy learning. When these three properties are combined, learning can diverge with the value estimates becoming unbounded. However, several algorithms successfully combine these three properties, which indicates that there is at least a partial gap in our understanding. In this work, we investigate the impact of the deadly triad in practice, in the context of a family of popular deep reinforcement learning models - deep Q-networks trained with experience replay - analyzing how the components of this system play a role in the emergence of the deadly triad, and in the agent’s performance

7.1.2. Multi-armed Bandit Theory

Corrupt Bandits for Preserving Local Privacy, [30]

We study a variant of the stochastic multi-armed bandit (MAB) problem in which the rewards are corrupted. In this framework, motivated by privacy preservation in online recommender systems, the goal is to maximize the sum of the (unobserved) rewards, based on the observation of transformation of these rewards through a stochastic corruption process with known parameters. We provide a lower bound on the expected regret of any bandit algorithm in this corrupted setting. We devise a frequentist algorithm, KLUCB-CF, and a Bayesian algorithm, TS-CF and give upper bounds on their regret. We also provide the appropriate corruption parameters to guarantee a desired level of local privacy and analyze how this impacts the regret. Finally, we present some experimental results that confirm our analysis.

A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption, [21] We study the problem of optimizing a function under a budgeted number of evaluations. We only assume that the function is locally smooth around one of its global optima. The difficulty of optimization is measured in terms of 1) the amount of noise b of the function evaluation and 2) the local smoothness, d, of the function. A smaller d results in smaller optimization error. We come with a new, simple, and parameter-free approach. First, for all values of b and d, this approach recovers at least the state-of-the-art regret guarantees. Second, our approach additionally obtains these results while being agnostic to the values of both b and d. This leads to the first algorithm that naturally adapts to an unknown range of noise b and leads to significant improvements in a moderate and low-noise regime. Third, our approach also obtains a remarkable improvement over the state-of-the-art SOO algorithm when the noise is very low which includes the case of optimization under deterministic feedback ($b = 0$). There, under our minimal local smoothness assumption, this improvement is of exponential magnitude and holds for a class of functions that covers the vast majority of functions that practitioners optimize ($d = 0$). We show that our algorithmic improvement is also borne out in the numerical experiments, where we empirically show faster convergence on common benchmark functions.

Best of both worlds: Stochastic & adversarial best-arm identification, [18]

We study bandit best-arm identification with arbitrary and potentially adversarial rewards. A simple random uniform learner obtains the optimal rate of error in the adversarial scenario. However, this type of strategy is suboptimal when the rewards are sampled stochastically. Therefore, we ask: Can we design a learner that performs optimally in both the stochastic and adversarial problems while not being aware of the nature of the rewards? First, we show that designing such a learner is impossible in general. In particular, to be robust to adversarial rewards, we can only guarantee optimal rates of error on a subset of the stochastic problems. We give a lower bound that characterizes the optimal rate in stochastic problems if the strategy is constrained to be robust to adversarial rewards. Finally, we design a simple parameter-free algorithm and show that its probability of error matches (up to log factors) the lower bound in stochastic problems, and it is also robust to adversarial ones.

Optimistic optimization of a Brownian, [31]

We address the problem of optimizing a Brownian motion. We consider a (random) realization W of a Brownian motion with input space in $[0, 1]$. Given W, our goal is to return an ϵ-approximation of its maximum using the smallest possible number of function evaluations, the sample complexity of the algorithm. We provide an algorithm with sample complexity of order $\log 2$ $(1/\epsilon)$. This improves over previous results of Al-Mharmah and Calvin (1996) and Calvin et al. (2017) which provided only polynomial rates. Our algorithm is adaptive-each query depends on previous values-and is an instance of the optimism-in-the-face-of-uncertainty principle.

Rotting bandits are no harder than stochastic ones, [37]

In bandits, arms' distributions are stationary. This is often violated in practice, where rewards change over time. In applications as recommendation systems, online advertising, and crowdsourcing, the changes may be triggered by the pulls, so that the arms' rewards change as a function of the number of pulls. In this paper, we consider the specific case of non-parametric rotting bandits, where the expected reward of an arm may decrease every time it is pulled. We introduce the filtering on expanding window average (FEWA) algorithm that at each round constructs moving averages of increasing windows to identify arms that are more likely to return high rewards when pulled once more. We prove that, without any knowledge on the decreasing behavior of the arms, FEWA achieves similar anytime problem-dependent, $\widetilde{\mathcal{O}}(\log(KT))$, and problem-independent, $\widetilde{\mathcal{O}}(\sqrt{KT})$, regret bounds of near-optimal stochastic algorithms as UCB1 of Auer et al. (2002a). This result substantially improves the prior result of Levine et al. (2017) which needed knowledge of the horizon and decaying parameters to achieve problem-independent bound of only $\widetilde{\mathcal{O}}(K^{1/3}T^{2/3})$. Finally, we report simulations confirming the theoretical improvements of FEWA.

Adaptive black-box optimization got easier: HCT only needs local smoothness, [41]

Hierarchical bandits is an approach for global optimization of extremely irregular functions. This paper provides new elements regarding POO, an adaptive meta-algorithm that does not require the knowledge of local smoothness of the target function. We first highlight the fact that the subroutine algorithm used in POO should have a small regret under the assumption of local smoothness with respect to the chosen partitioning, which is unknown if it is satisfied by the standard subroutine HOO. In this work, we establish such regret guarantee for HCT, which is another hierarchical optimistic optimization algorithm that needs to know the smoothness. This confirms the validity of POO. We show that POO can be used with HCT as a subroutine with a regret upper bound that matches the one of best-known algorithms using the knowledge of smoothness up to a $\sqrt{}$ log n factor.

Boundary Crossing Probabilities for General Exponential Families, [16]

Multi-Player Bandits Revisited, [22]

Multi-player Multi-Armed Bandits (MAB) have been extensively studied in the literature, motivated by applications to Cognitive Radio systems. Driven by such applications as well, we motivate the introduction of several levels of feedback for multi-player MAB algorithms. Most existing work assume that sensing information is available to the algorithm. Under this assumption, we improve the state-of-the-art lower bound for the regret of any decentralized algorithms and introduce two algorithms, RandTopM and MCTopM, that are shown to empirically outperform existing algorithms. Moreover, we provide strong theoretical guarantees for these algorithms, including a notion of asymptotic optimality in terms of the number of selections of bad arms. We then introduce a promising heuristic, called Selfish, that can operate without sensing information, which is crucial for emerging applications to Internet of Things networks. We investigate the empirical performance of this algorithm and provide some first theoretical elements for the understanding of its behavior.

Pure Exploration in Infinitely-Armed Bandit Models with Fixed-Confidence, [19]

We consider the problem of near-optimal arm identification in the fixed confidence setting of the infinitely armed bandit problem when nothing is known about the arm reservoir distribution. We (1) introduce a PAC-like framework within which to derive and cast results; (2) derive a sample complexity lower bound for near-optimal arm identification; (3) propose an algorithm that identifies a nearly-optimal arm with high probability and derive an upper bound on its sample complexity which is within a log factor of our lower bound; and (4) discuss whether our $\log^2(1/\delta)$ dependence is inescapable for "two-phase" (select arms first, identify the best later) algorithms in the infinite setting. This work permits the application of bandit models to a broader class of problems where fewer assumptions hold.

Aggregation of Multi-Armed Bandits Learning Algorithms for Opportunistic Spectrum Access, [23]

Multi-armed bandit algorithms have been recently studied and evaluated for Cognitive Radio (CR), especially in the context of Opportunistic Spectrum Access (OSA). Several solutions have been explored based on various models, but it is hard to exactly predict which could be the best for real-world conditions at every instants. Hence, expert aggregation algorithms can be useful to select on the run the best algorithm for a specific situation. Aggregation algorithms, such as Exp4 dating back from 2002, have never been used for OSA learning, and we show that it appears empirically sub-efficient when applied to simple stochastic problems. In this article, we present an improved variant, called Aggregator. For synthetic OSA problems modeled as Multi-Armed Bandit (MAB) problems, simulation results are presented to demonstrate its empirical efficiency. We combine classical algorithms, such as Thompson sampling, Upper-Confidence Bounds algorithms (UCB and variants), and Bayesian or Kullback-Leibler UCB. Our algorithm offers good performance compared to state-of-the-art algorithms (Exp4, CORRAL or LearnExp), and appears as a robust approach to select on the run the best algorithm for any stochastic MAB problem, being more realistic to real-world radio settings than any tuning-based approach.

What Doubling Tricks Can and Can't Do for Multi-Armed Bandits, [47]

An online reinforcement learning algorithm is anytime if it does not need to know in advance the horizon T of the experiment. A well-known technique to obtain an anytime algorithm from any non-anytime algorithm is the "Doubling Trick". In the context of adversarial or stochastic multi-armed bandits, the performance of an algorithm is measured by its regret, and we study two families of sequences of growing horizons (geometric and exponential) to generalize previously known results that certain doubling tricks can be used to conserve certain regret bounds. In a broad setting, we prove that a geometric doubling trick can be used to conserve (minimax) bounds in $R\_T = O(\sqrt{T})$ but cannot conserve (distribution-dependent) bounds in

$R\_T = O(\log T)$. We give insights as to why exponential doubling tricks may be better, as they conserve bounds in $R\_T = O(\log T)$, and are close to conserving bounds in $R\_T = O(\sqrt{T})$.

Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals, [50]

This paper presents new deviation inequalities that are valid uniformly in time under adaptive sampling in a multi-armed bandit model. The deviations are measured using the Kullback-Leibler divergence in a given one-dimensional exponential family, and may take into account several arms at a time. They are obtained by constructing for each arm a mixture martingale based on a hierarchical prior, and by multiplying those martingales. Our deviation inequalities allow us to analyze stopping rules based on generalized likelihood ratios for a large class of sequential identification problems, and to construct tight confidence intervals for some functions of the means of the arms.

7.1.3. Stochastic Games

Actor-Critic Fictitious Play in Simultaneous Move Multistage Games, [36]

Fictitious play is a game theoretic iterative procedure meant to learn an equilibrium in normal form games. However, this algorithm requires that each player has full knowledge of other players' strategies. Using an architecture inspired by actor-critic algorithms, we build a stochastic approximation of the fictitious play process. This procedure is on-line, decentralized (an agent has no information of others' strategies and rewards) and applies to multistage games (a generalization of normal form games). In addition, we prove convergence of our method towards a Nash equilibrium in both the cases of zero-sum two-player multistage games and cooperative multistage games. We also provide empirical evidence of the soundness of our approach on the game of Alesia with and without function approximation.

Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling, [39]

Learning the minimum/maximum mean among a finite set of distributions is a fundamental sub-task in planning, game tree search and reinforcement learning. We formalize this learning task as the problem of sequentially testing how the minimum mean among a finite set of distributions compares to a given threshold. We develop refined non-asymptotic lower bounds, which show that optimality mandates very different sampling behavior for a low vs high true minimum. We show that Thompson Sampling and the intuitive Lower Confidence Bounds policy each nail only one of these cases. We develop a novel approach that we call Murphy Sampling. Even though it entertains exclusively low true minima, we prove that MS is optimal for both possibilities. We then design advanced self-normalized deviation inequalities, fueling more aggressive stopping rules. We complement our theoretical guarantees by experiments showing that MS works best in practice.

7.1.4. Online Kernel and Graph-Based Methods

Improved large-scale graph learning through ridge spectral sparsification, [25]

The representation and learning benefits of methods based on graph Laplacians, such as Laplacian smoothing or harmonic function solution for semi-supervised learning (SSL), are empirically and theoretically well supported. Nonetheless , the exact versions of these methods scale poorly with the number of nodes n of the graph. In this paper, we combine a spectral sparsification routine with Laplacian learning. Given a graph G as input, our algorithm computes a sparsifier in a distributed way in $O(n \log 3(n))$ time, $O(m \log 3(n))$ work and $O(n \log(n))$ memory, using only $\log(n)$ rounds of communication. Furthermore , motivated by the regularization often employed in learning algorithms, we show that constructing sparsifiers that preserve the spectrum of the Laplacian only up to the regularization level may drastically reduce the size of the final graph. By constructing a spectrally-similar graph, we are able to bound the error induced by the sparsification for a variety of downstream tasks (e.g., SSL). We empirically validate the theoretical guarantees on Amazon co-purchase graph and compare to the state-of-the-art heuristics.

DPPy: Sampling Determinantal Point Processes with Python, [49] Determinantal point processes (DPPs) are specific probability distributions over clouds of points that are used as models and computational tools across physics, probability, statistics, and more recently machine learning. Sampling from DPPs is a challenge and therefore we present DPPy, a Python toolbox that gathers known exact and approximate sampling algorithms. The project is hosted on GitHub and equipped with an extensive documentation. This documentation takes the form of a short survey of DPPs and relates each mathematical property with DPPy objects.

Streaming kernel regression with provably adaptive mean, variance, and regularization, [14]

We consider the problem of streaming kernel regression, when the observations arrive sequentially and the goal is to recover the underlying mean function, assumed to belong to an RKHS. The variance of the noise is not assumed to be known. In this context, we tackle the problem of tuning the regularization parameter adaptively at each time step, while maintaining tight confidence bounds estimates on the value of the mean function at each point. To this end, we first generalize existing results for finite-dimensional linear regression with fixed regularization and known variance to the kernel setup with a regularization parameter allowed to be a measurable function of past observations. Then, using appropriate self-normalized inequalities we build upper and lower bound estimates for the variance, leading to Bersntein-like concentration bounds. The later is used in order to define the adaptive regularization. The bounds resulting from our technique are valid uniformly over all observation points and all time steps, and are compared against the literature with numerical experiments. Finally, the potential of these tools is illustrated by an application to kernelized bandits, where we revisit the Kernel UCB and Kernel Thompson Sampling procedures, and show the benefits of the novel adaptive kernel tuning strategy.

7.2. Applications

7.2.1. Dialogue Systems and Natural Language

FiLM: Visual Reasoning with a General Conditioning Layer, [35]

We introduce a general-purpose conditioning method for neural networks called FiLM: Feature-wise Linear Modulation. FiLM layers influence neural network computation via a simple , feature-wise affine transformation based on conditioning information. We show that FiLM layers are highly effective for visual reasoning - answering image-related questions which require a multi-step, high-level process - a task which has proven difficult for standard deep learning methods that do not explicitly model reasoning. Specifically, we show on visual reasoning tasks that FiLM layers 1) halve state-of-the-art error for the CLEVR benchmark, 2) modulate features in a coherent manner, 3) are robust to ablations and architectural modifications, and 4) generalize well to challenging, new data from few examples or even zero-shot.

End-to-End Automatic Speech Translation of Audiobooks, [24]

We investigate end-to-end speech-to-text translation on a corpus of audiobooks specifically augmented for this task. Previous works investigated the extreme case where source language transcription is not available during learning nor decoding , but we also study a midway case where source language transcription is available at training time only. In this case, a single model is trained to decode source speech into target text in a single pass. Experimental results show that it is possible to train compact and efficient end-to-end speech translation models in this setup. We also distribute the corpus and hope that our speech translation baseline on this corpus will be challenged in the future.

Visual Reasoning with Multi-hop Feature Modulation, [42]

Recent breakthroughs in computer vision and natural language processing have spurred interest in challenging multi-modal tasks such as visual question-answering and visual dialogue. For such tasks, one successful approach is to condition image-based convolutional network computation on language via Feature-wise Linear Modulation (FiLM) layers, i.e., per-channel scaling and shifting. We propose to generate the parameters of FiLM layers going up the hierarchy of a convolutional network in a multi-hop fashion rather than all at once, as in prior work. By alternating between attending to the language input and generating FiLM layer parameters, this approach is better able to scale to settings with longer input sequences such as dialogue. We demonstrate that multi-hop FiLM generation achieves state-of-the-art for the short input sequence task ReferIt-on-par with single-hop FiLM generation-while also significantly outperforming prior state-of-the-art and single-hop FiLM generation on the GuessWhat?! visual dialogue task.

7.2.2. Recommendation systems

Recurrent Neural Networks for Long and Short-Term Sequential Recommendation, [54]

Recommender systems objectives can be broadly characterized as modeling user preferences over short- or long-term time horizon. A large body of previous research studied long-term recommendation through dimensionality reduction techniques applied to the historical user-item interactions. A recently introduced session-based recommendation setting highlighted the importance of modeling short-term user preferences. In this task, Recurrent Neural Networks (RNN) have shown to be successful at capturing the nuances of user's interactions within a short time window. In this paper, we evaluate RNN-based models on both short-term and long-term recommendation tasks. Our experimental results suggest that RNNs are capable of predicting immediate as well as distant user interactions. We also find the best performing configuration to be a stacked RNN with layer normalization and tied item embeddings.

Fighting Boredom in Recommender Systems with Linear Reinforcement Learning, [43]

A common assumption in recommender systems (RS) is the existence of a best fixed recommendation strategy. Such strategy may be simple and work at the item level (e.g., in multi-armed bandit it is assumed one best fixed arm/item exists) or implement more sophisticated RS (e.g., the objective of A/B testing is to find the best fixed RS and execute it thereafter). We argue that this assumption is rarely verified in practice, as the recommendation process itself may impact the user's preferences. For instance, a user may get bored by a strategy, while she may gain interest again, if enough time passed since the last time that strategy was used. In this case, a better approach consists in alternating different solutions at the right frequency to fully exploit their potential. In this paper, we first cast the problem as a Markov decision process, where the rewards are a linear function of the recent history of actions, and we show that a policy considering the long-term influence of the recommendations may outperform both fixed-action and contextual greedy policies. We then introduce an extension of the UCRL algorithm (LINUCRL) to effectively balance exploration and exploitation in an unknown environment, and we derive a regret bound that is independent of the number of states. Finally, we empirically validate the model assumptions and the algorithm in a number of realistic scenarios.

7.2.3. Autonomous car

A Survey of State-Action Representations for Autonomous Driving, [51]

Approximate Robust Control of Uncertain Dynamical Systems, [40]

This work studies the design of safe control policies for large-scale non-linear systems operating in uncertain environments. In such a case, the robust control framework is a principled approach to safety that aims to maximize the worst-case performance of a system. However, the resulting optimization problem is generally intractable for non-linear systems with continuous states. To overcome this issue, we introduce two tractable methods that are based either on sampling or on a conservative approximation of the robust objective. The proposed approaches are applied to the problem of autonomous driving.

7.2.4. Software development

Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation, [12]

Can the execution of a software be perturbed without breaking the correctness of the output? In this paper, we devise a novel protocol to answer this rarely investigated question. In an experimental study, we observe that many perturbations do not break the correctness in ten subject programs. We call this phenomenon "correctness attraction". The uniqueness of this protocol is that it considers a systematic exploration of the perturbation space as well as perfect oracles to determine the correctness of the output. To this extent, our findings on the stability of software under execution perturbations have a level of validity that has never been reported before in the scarce related work. A qualitative manual analysis enables us to set up the first taxonomy ever of the reasons behind correctness attraction.

This paper has attracted a significant interest in the SE community. This work has been invited for an oral presentation (along a 1 page summary) at the 40th International Conference on Software Engineering, the main conference in software engineering. It has then been invited on the IEEE Software review blog.

SMPyBandits: an Experimental Framework for Single and Multi-Players Multi-Arms Bandits Algorithms in Python, [46]

SMPyBandits is a package for numerical simulations on single-player and multi-players Multi-Armed Bandits (MAB) algorithms, written in Python (2 or 3). This library is the most complete open-source implementation of state-of-the-art algorithms tackling various kinds of sequential learning problems referred to as Multi-Armed Bandits. It is extensive, simple to use and maintain, with a clean and well documented codebase. It allows fast prototyping of experiments, with an easy configuration system and command-line options to customize experiments.

Lilian Besson developed a library for multi-armed bandit algorithms in Python for single and multi-player bandits.

7.2.5. Deep Learning

FiLM: Visual Reasoning with a General Conditioning Layer, [35]

We introduce a general-purpose conditioning method for neural networks called FiLM: Feature-wise Linear Modulation. FiLM layers influence neural network computation via a simple , feature-wise affine transformation based on conditioning information. We show that FiLM layers are highly effective for visual reasoning - answering image-related questions which require a multi-step, high-level process - a task which has proven difficult for standard deep learning methods that do not explicitly model reasoning. Specifically, we show on visual reasoning tasks that FiLM layers 1) halve state-of-the-art error for the CLEVR benchmark, 2) modulate features in a coherent manner, 3) are robust to ablations and architectural modifications, and 4) generalize well to challenging, new data from few examples or even zero-shot.

Feature-wise transformations, [13]

i-RevNet: Deep Invertible Networks, [32]

It is widely believed that the success of deep convolutional networks is based on progressively discarding uninformative variability about the input with respect to the problem at hand. This is supported empirically by the difficulty of recovering images from their hidden representations, in most commonly used network architectures. In this paper we show via a one-to-one mapping that this loss of information is not a necessary condition to learn representations that generalize well on complicated problems, such as ImageNet. Via a cascade of homeomorphic layers, we build the i-RevNet, a network that can be fully inverted up to the final projection onto the classes, i.e. no information is discarded. Building an invertible architecture is difficult, for one, because the local inversion is ill-conditioned, we overcome this by providing an explicit inverse. An analysis of i-RevNets learned representations suggests an alternative explanation for the success of deep networks by a progressive contraction and linear separation with depth. To shed light on the nature of the model learned by the i-RevNet we reconstruct linear interpolations between natural image representations.

Compressing the Input for CNNs with the First-Order Scattering Transform, [33]

We study the first-order scattering transform as a candidate for reducing the signal processed by a convolutional neural network (CNN). We study this transformation and show theoretical and empirical evidence that in the case of natural images and sufficiently small translation invariance, this transform preserves most of the signal information needed for classification while substantially reducing the spatial resolution and total signal size. We show that cascading a CNN with this representation performs on par with ImageNet classification models commonly used in downstream tasks such as the ResNet-50. We subsequently apply our trained hybrid ImageNet model as a base model on a detection system, which has typically larger image inputs. On Pascal VOC and COCO detection tasks we deliver substantial improvements in the inference speed and training memory consumption compared to models trained directly on the input image.

Visual Reasoning with Multi-hop Feature Modulation, [42] Recent breakthroughs in computer vision and natural language processing have spurred interest in challenging multi-modal tasks such as visual question-answering and visual dialogue. For such tasks, one successful approach is to condition image-based convolutional network computation on language via Feature-wise Linear Modulation (FiLM) layers, i.e., per-channel scaling and shifting. We propose to generate the parameters of FiLM layers going up the hierarchy of a convolutional network in a multi-hop fashion rather than all at once, as in prior work. By alternating between attending to the language input and generating FiLM layer parameters, this approach is better able to scale to settings with longer input sequences such as dialogue. We demonstrate that multi-hop FiLM generation achieves state-of-the-art for the short input sequence task ReferIt-on-par with single-hop FiLM generation-while also significantly outperforming prior state-of-the-art and single-hop FiLM generation on the GuessWhat?! visual dialogue task.

8. Bilateral Contracts and Grants with Industry

8.1. Bilateral Contracts with Industry

8.1.1. Lelivrescolaire.fr

- contract with http://Lelivrescolaire.fr; PI: Michal Valko

  Title: Sequential Machine Learning for Adaptive Educational Systems

  Duration: Mar. 2018 – Feb. 2021

  Abstract: Adaptive educational content are technologies which adapt to the difficulties encountered by students. With the rise of digital content in schools, the mass of data coming from education enables but also ask for machine learning methods. Since 2010, Lelivrescolaire.fr has been developing some learning materials for teachers and students through collaborative creation process. For instance, during the school year 2015/2016, students has achieved more than 8 000 000 exercises on its homework platform Afterclasse.fr. Our approach would be based on sequential machine learning: the algorithm learns to recommend some exercises which adapt to students gradually as they answer.
  Participants: Julien Seznec, Michal Valko.

8.1.2. Sidexa

- contract with “Sidexa”; PI: Philippe Preux

  Title: vision applied to the segmentation and recognition of cars and car related documents.

  Duration: 6 months

  Abstract: this is a follow-up to the successful contract realized in 2017 with Sidexa. We studied multi-class supervised classification problems in order to classify documents related to a car, and also to identify various characteristics of a car, such as its color, its make, its type.

  This work is done with an InriaTech engineer.
  Participant: Philippe Preux.

8.1.3. Renault

- contract with Renault; PI: Philippe Preux

  Title: Control of an autonomous vehicle

  Duration: 3 years (12/2017–11/2020)

  Abstract: This contract comes along the CIFRE grant on the same topic. This work is done in collaboration with the NON-A team-project.
  Participants: Édouard Leurent, Odalric Maillard, Philippe Preux.

8.1.4. Critéo

- contract with “Criteo”; PI: Philippe Preux

  Title: Computational advertizing

  Duration: 3 years (12/2017–11/2020)

  Abstract: This contract comes along the CIFRE grant on the same topic. The goal is to investigate reinforcement learning and deep learning on the problem of ad selection on the Internet.
  Participants: Philippe Preux, Kiewan Villatel.

8.1.5. Orange Labs

- contract with “Orange Labs”; PI: Olivier Pietquin

  Title: Inter User Transfer in dialogue systems

  Duration: 3 years

  Abstract: This contract comes along the CIFRE grant on the same topic. The research aims at developing new algorithms to learn fast adaptation strategies for dialogue systems when a new user starts using them while we collected data from previous interactions with other users. Especially, it addresses the cold-start problem encountered when a new user faces the system, before samples can be collected to optimize the interaction strategy.
  Participants: Merwan Barlier, Nicolas Carrara, Olivier Pietquin.

8.1.6. 55

- contract with “55”; PI: Jérémie Mary

  Title: Novel Learning and Exploration-Exploitation Methods for Effective Recommender Systems

  Duration: Oct. 2015 – Sep. 2018

  Abstract: This contract comes along the CIFRE grant on the same topic. 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.

  The PhD was defended in Fall 2018.
  Participants: Jérémie Mary, Romain Warlop.

8.1.7. AB-Tasty

- Thompson Sampling for A/B/C Testing with Delayed Conversions; PI: Émilie Kaufmann

  Duration: 1 month

  Abstract: We investigated the use of Thompson Sampling as well as other state-of-the-art methods for the stochastic MAB problem in the context of delayed feedback. We provided theoretical justification for a method developed by AB Tasty, and proposed some variants of it, as well as a comparison with existing methods from the literature.
  Participant: Émilie Kaufmann.

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 (Rémi 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 favorably 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, Émilie Kaufmann.

- *Title*: BAnDits for non-Stationarity and Structure
- *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: Jérémie Mary, Michal Valko.

- *Title*: Extraction and Transfer of Knowledge in Reinforcement Learning
- *Type*: National Research Agency (ANR-9011)
- *PI*: M. Valko
- *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.

9.1.4. Grant of Fondation Mathématique Jacques Hadamard

Participants: Michal Valko, Matteo Pirotta, Alessandro Lazaric, Ronan Fruit.

- *Title*: Theoretically grounded efficient algorithms for high-dimensional and continuous reinforcement learning
- *Type*: PGMO-IRMO, funded by Criteo
- *PI*: M. Valko
- *Criteo contact*: Marc Abeille
- *Duration*: 2018-2020
- *Abstract*: While learning how to behave optimally in an unknown environment, a reinforcement learning (RL) agent must trade off the exploration needed to collect new information about the dynamics and reward of the environment, and the exploitation of the experience gathered so far to gain as much reward as possible. A good measure of the agent's performance is the regret, which measures the difference between the performance of optimal policy and the actual rewards accumulated by the agent. Two common approaches to the exploration-exploitation dilemma with provably good regret guarantees are the optimism in the face of uncertainty principle and Thompson Sampling. While these approaches have been successfully applied to small environments with a finite number of states and action (tabular scenario), existing approach for large or continuous environments either rely on heuristics and come with no regret guarantees, or can be proved to achieve small regret but cannot be implemented efficiently. In this project, we propose to make a significant contribution in the understanding of large and/or continuous RL problems by developing and analyzing new algorithms that perform well both in theory and practice. This research line can have a practical impact in all the applications requiring continuous interaction with an unknown environment. Recommendation systems belong to this category and, by definition, they can be modeled has a sequence of repeated interaction between a learning agent and a large (possibly continuous) environment.

9.1.5. National Partners

- ENS Paris-Saclay
  - M. Valko collaborated with V. Perchet on structured bandit problem. They co-supervise a PhD student (P. Perrault) together.
- Institut de Mathématiques de Toulouse, then Ecole Normale Supérieure de Lyon
  - E. Kaufmann collaborated with Aurélien Garivier on sequential testing and structured bandit problems.
- Centrale-Supélec Rennes
  - E. Kaufmann co-advises Lilian Besson, who works at CentraleSupélec with Christophe Moy on MAB for cognitive radio and Internet-of-Things communications.
- Participation to the Inria Project Lab (IPL) “HPC – Big Data”. Started in 2018, this IPL gathers a dozen Inria team-projects, mixing researchers in HPC with researchers in machine learning and data science. SEQUEL contribution in this project is about how we can take advantage of HPC for our computational needs regarding deep learning and deep reinforcement learning, and also how such learning algorithms might be redesigned or re-implemented in order to take advantage of HPC architectures.

9.2. European Initiatives

9.2.1. FP7 & H2020 Projects

9.2.1.1. H2020 BabyRobot

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 (mind reading) 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 education 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-in hand 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.

9.2.1.2. CHIST-ERA DELTA

Participants: Michal Valko, Émilie Kaufmann.

Program: CHIST-ERA

Project acronym: DELTA

Project title: Dynamically Evolving Long-Term Autonomy

Duration: October 2017 - December 2021

Coordinator: Anders Jonsson (PI)

Inria Coordinator: Michal Valko

Other partners: UPF Spain, MUL Austria, ULG Belgium

Abstract: Many complex autonomous systems (e.g., electrical distribution networks) repeatedly select actions with the aim of achieving a given objective. Reinforcement learning (RL) offers a powerful framework for acquiring adaptive behavior in this setting, associating a scalar reward with each action and learning from experience which action to select to maximise long-term reward. Although RL has produced impressive results recently (e.g., achieving human-level play in Atari games and beating the human world champion in the board game Go), most existing solutions only work under strong assumptions: the environment model is stationary, the objective is fixed, and trials end once the objective is met. The aim of this project is to advance the state of the art of fundamental research in lifelong RL by developing several novel RL algorithms that relax the above assumptions. The new algorithms should be robust to environmental changes, both in terms of the observations that the system can make and the actions that the system can perform. Moreover, the algorithms should be able to operate over long periods of time while achieving different objectives. The proposed algorithms will address three key problems related to lifelong RL: planning, exploration, and task decomposition. Planning is the problem of computing an action selection strategy given a (possibly partial) model of the task at hand. Exploration is the problem of selecting actions with the aim of mapping out the environment rather than achieving a particular objective. Task decomposition is the problem of defining different objectives and assigning a separate action selection strategy to each. The algorithms will be evaluated in two realistic scenarios: active network management for electrical distribution networks, and microgrid management. A test protocol will be developed to evaluate each individual algorithm, as well as their combinations.

9.2.1.3. CHIST-ERA IGLU

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. Allocate

Participants: Pierre Perrault, Julien Seznec, Michal Valko, Émilie Kaufmann, Odalric Maillard.

Title: Adaptive allocation of resources for recommender systems

Inria contact: Michal Valko

International Partner (Institution - Laboratory - Researcher):

Otto-von-Guericke-Universität Magdeburg A. Carpentier

Start year: 2017

We plan to improve a practical scenario of *resource allocation in market surveys*, such as product appraisals and music recommendation. In practice, the market is typically divided into segments: geographic regions, age groups, ...These groups are then queried for preference with some fixed rule of a number of queries per group. This testing is *costly and non-adaptive*. The reason is some groups are easier to estimate than others, but this is impossible to know a priori. Our challenge is **adaptively allocate the optimal number of samples** to each group and improve the efficient of market studies, by providing *sample-efficient* solutions. In 2018 we made big advances that resulted in two new research results, currently under review.

9.3.2. Inria International Partners

9.3.2.1. Declared Inria International Partners

SequeL

Title: The multi-armed bandit problem

International Partner (Institution - Laboratory - Researcher):

University of Leoben (Austria) Peter Auer

Duration: 2014 - 2018

Start year: 2014

In a nutshell, the collaboration is focusing on nonparametric algorithms for active learning problems, mainly involving theoretical analysis of reinforcement learning and bandits problems beyond the traditional settings of finite-state MDPs (for RL) or i.i.d. rewards (for bandits). Peter Auer from University of Leoben is a worldwide leader in the field, having introduced the UCB approach around 2000, along with its finite-time analysis. Today, SequeL is likely to be the largest research group working in this field in the world, enjoying worldwide recognition. SequeL and P. Auer's group have been collaborating for a couple of years now; they have co-authored papers, visited each other (sabbatical stay, post-doc), coorganized workshops; the STREP Complacs partially funds this very active collaboration.

9.3.2.2. CWI

We also collaborate with P. Grunwald, and W. Koolen through the associate team headed by Benjamin Guedj from Modal.

9.3.3. Participation in Other International Programs

In 2017, we mentioned many collaborations with: Adobe, MIT, Stanford, Leoben, ...

Massachusetts Institute of Technology

Victor-Emmanuel Brunel *Collaborator*

M. Valko collaborated with V.-E. Brunel on the estimation of low rank determinantal point processes useful for diverse recommender systems.

Otto-von-Guericke-Universität Magdeburg

Alexandra Carpentier *Collaborator*

M. Valko collaborated with A. Carpentier on adaptive estimation of the block-diagonal matrices with application to market segmentations. This collaboration formalized in September 2017 by creating a north-european associate team. which results in two finished results.

Adobe Research

Y. Abbasi-Yadkori *Collaborator*

M. Valko collaborated on learning in unpredictable but potentially easy environment. This led to a publication in COLT 2018.

University of California, Berkeley

Peter Bartlett *Collaborator*

Victor Gabillon *Collaborator*

Alain Malek *Collaborator*

M. Valko collaborated with P.Barlett, V. Gabillon, and A. Malek on the sample complexities in unknown type of environments.

DeepMind London

Rémi Munos *Collaborator*

M. Valko collaborated with R. Munos on Brownian motion maximization, important for stock value predictions. This led to a publication in NIPS 2018.

Mila, Université de Montréal

A. Courville *Collaborator*

A. Touati *Collaborator*

F. Strub and O. Pietquin collaborate on deep reinforcement learning for language acquisition. This led to several papers at IJCAI, CVPR, and NIPS, as well as the Guesswhat?! dataset and protocol, and the HOME dataset.

M. Valko collaborates on faster learning in submodular learning with limited feedback. This setting has application in marketing when we want to select the inventory while maximizing the profit.

McGill University, Montreal

A. Durand, J. Pineau *Collaborator*

A. Durand and OA. Maillard collaborate on a project of structured bandits, with application in physics (calibration).

Northeastearn University, Boston

M. Aziz, J. Anderton, J. Aslam *Collaborator*

E. Kaufmann collaborate with M. Aziz, J. Anderton and J. Aslam on a project on infinite bandits, which led to an ALT 2018 publication. E. Kaufmann also collaborates with M. Aziz on bandits for phase I clinical trials. This led to the submission of a paper to the Biometrics journal.

9.4. International Research Visitors

9.4.1. Visits of International Scientists

- Xiotian Yu, 1 week, the Chinese University of Hong-Kong
- Junpei Komiyama, 6 weeks, Tokyo University
- Abbas Mehrabian, 1 week, McGill University
- Audrey Durand, 2 weeks, McGill University
- Andrea Locatelli, 2 weeks, Otto-von-Guericke-Universität Magdeburg, Germany
- Jill-Jênn Vie, 1 week, RIKEN AIP, Tokyo, Japan
- Peter Grünwald, 2 times two days (8 hour lectures), CWI and Leiden University, Amsterdam, Netherlands
- Wouter Koolen, 1 week, CWI, Amsterdam, Netherlands

9.4.1.1. Internships

- Quentin Burthier, ENSTA ParisTech, from Jun 2018 until Aug 2018
- Edouard Dendauw, from May 2018 until Jul 2018
- Thibault Felicite, Jul 2018

- Robert Lindland, MIT, from May 2018 until Aug 2018
- Jian Qian, ENS, from May 2018 until Oct 2018
- Hassan Saber, Centrale Paris, from Apr 2018 until Aug 2018
- Benoit Schmitt, Centrale Nantes, from Mar 2018 until Aug 2018
- Han Shao, PhD student from the Chinese University in Hong-Kong, from Oct 2018 until Nov 2018
- Annie Yun, MIT, from May 2018 until Aug 2018
- Arnaud Fanthomme, ENS, from Apr 2018 until Aug 2018

9.4.2. Visits to International Teams

9.4.2.1. Other visits

- OA. Maillard: August, Visit of Aufrey Durand at Mc Gill University (2 weeks)
- OA. Maillard: Novembre, Invited visit of Junya Honda at Tokyo University (4 days)
- E. Kaufmann: March, April, Visit of Wouter Koolen at CWI, Amsterdam (2 times 1 week)

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. EWRL 2018

We organized the 14th European Workshop on Reinforcement Learning (EWRL) in October 2018 in Lille. 183 people registered. Despite its name, the audience goes really beyond Europe with 42 from North Amercia, 38 from France, 19 from Germany, 16 from the UK, 12 from Italy, 12 from Israel, 9 from Belgium, ... 40% of participants were students, mostly PhD students, but also some Master students. Among non students, 40% came from industry, the other 60% being academics. We had a quite unique panel of invited speakers highlighted by historical figures of reinforcement learning with Prof. Richard Sutton (U. Alberta and Deepmind), bandit theory with Tze Leung Lai (Stanford U.). EWRL is really the main scientific event on reinforcement learning in the world today.

After 2008 and 2015, this is the third time EWRL is organized in Lille.

10.1.1.2. General Chair, Scientific Chair

M. Valko was a program co-chair for CNRS Summer school on Networks, Graphs, and Machine Learning (RESCOM 2018)

10.1.1.3. Member of the Organizing Committees

- F. Strub, co-organizer of the workshop “Visually Grounded Interaction and Language (ViGIL)” at NIPS 2018
- M. Valko was an organizing co-chair of the ITS Workshop: Optimizing Human Learning (ITS 2018)
- R. Fruit, M. Seurin, M. Pirotta, F. Strub organized the 14th European Workshop on Reinforcement Learning

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

- Philippe Preux: SPC IJCAI 2018; PC of ICML, ECML, LOD, EWRL, and French speaking conferences: EGC, SFC JFPDA
- Michal Valko: Area Chair of NIPS 2018, Top 10 reviewer recognition for reviewing at ICML
- Emilie Kaufmann: PC Chair for WiML 2018, Top 10 reviewer recognition for reviewing at ICML
- Odalric-ambrym Maillard: PC chair for ALT

10.1.2.2. Reviewer

Members of SEQUEL have been involved in the following reviewing activities for conferences in 2018:

- AI&Stats, NIPS, ALT, ICML, COLT, IJCAI

10.1.3. Journal

10.1.3.1. Reviewer - Reviewing Activities

- The Annals of Statistics
- Journal of Machine Learning Research
- Machine Learning Research
- Bernoulli
- Annual Reviews in Control
- European Journal of Operation Research
- Information and Inference: a Journal of the IMA (Institute of Mathematics and its Application)
- Operations Research
- IEEE Transactions on Signal Processing

10.1.4. Invited Talks

- Odalric-Ambrym Maillard: invited Opponent for the PhD defence of Stefan Magureanu (Stockholm, Sweden), February 2018
- Odalric-Ambrym Maillard: invited speaker at LTCI, Telecom ParisTech, February 2018
- Odalric-Ambrym Maillard: invited speaker at Journées Probabilités et statistiques de Lille, June 22, 2018
- Odalric-Ambrym Maillard: invited speaker at RL Lab, McGill University, August 23, 2018
- Odalric-Ambrym Maillard: invited speaker at the 21st IBIS conference (Sapporo, Japan), November 06, 2018
- Odalric-Ambrym Maillard: invited speaker at the RIKEN Institute (Tokyo, Japan), November 07, 2018
- Michal Valko: *The power of graphs in speeding up online learning and decision making* Presented on October 23rd, DeepMind, London, UK (*DeepMind 2018*)
- Michal Valko: *Active block-matrix completion with adaptive confidence sets*, Presented on September 10–13th, 2018, International Workshop on Optimization and Machine Learning, CIMI, Toulouse (*CIMI 2018*)
- Michal Valko: *Online influence maximization*, Presented on May 14th, 2018, Workshop on Graph Learning, LINCS, Paris (*LINCS 2018*)
- Michal Valko: *Recommender systems*, Presented on March 22nd, 2018, Journée Big data, Polytech'Lille (*Polytech'Lille 2018*)
- Michal Valko: *Pliable rejection sampling*, Presented on February 8th, 2018 at GDR Isis, Télécom ParisTech in Paris (*ISIS 2018*)
- Michal Valko: *Graph Bandits*, Presented on January 7th, 2018 at MIST conference in Rajecká Lesná (*MIST 2018*)
- Pierre Perrault: *Stochastic multi-arm bandit problem and some extensions*, Presented on November, 23rd, 2018 at Lambda seminar at Université de Bordeaux (*Lambda 2018*)
- Emilie Kaufmann: *(Optimal) Best Arm Identification and applications to Monte-Carlo Tree Search*, presented on January, 18th, 2018 at the Probability and Statistics seminar of IECL, Nancy
- Emilie Kaufmann: *Bandits (for) Games*, presented on March 26th, 2018 at Amazon Research, Berlin
- Emilie Kaufmann, *Bandits (for) Games*, presented on April 25th, 2018 as an invited talk to the Workshop on Modern Challenges on Learning Theory at Univerisé de Montréal

- Emilie Kaufmann, *Bandits (for) Games*, presented on June 13th, 2018 as an invited talk to the Paris Symposium on Game Theory, Paris
- Emilie Kaufmann *New tools for Adaptive Testing and Applications to Bandit Problems*, presented on December 3rd, 2018 at the Probability and Statistics seminar of IRMA, Strasbourg

10.1.5. Scientific Expertise

- Philippe Preux was a member of the hiring committee for CR at Inria Nancy
- Philippe Preux was a member of the hiring committee for an associate professor at Université de Lille
- Philippe Preux evaluated submissions to ANRT (he also declined many such invitations due to lack of time)
- Philippe Preux was a member of an auditing committee of an international company which can not be named (NDA)
- Philippe Preux participates to a "AI mission" with an (other) international company which can not be named (NDA)
- Odalric-Ambrym Maillard evaluated a submission to OTKA (Hungarian ANR), and to ANR.
- M. Valko is an elected member of the evaluation committee and participates in the hiring, promotion, and evaluation juries of Inria, notably
  - Selection committee for Inria award for scientific excellence of confirmed researchers
  - Inria working group for the creation of team RandOpt
  - National committee for the secondments at Inria
- Michal Valko participates in a collaboration with an international company which can not be named (NDA)
- Emilie Kaufmann was a member of the hiring committee for an associate professor position at Université de Lille
- Emilie Kaufmann was a member of the hiring committee for an associate professor at ENS Paris (Computer Science departement)

10.1.6. Research Administration

- Philippe Preux is:
  - "délégué scientifique adjoint" of the Inria center in Lille
  - member of the Inria evaluation committee (CE)
  - member of the Inria internal scientific committee (COSI)
  - member of the scientific committee of CRIStAL
  - the head of the "Data Intelligence" thematic group at CRIStAL
- Michal Valko is a member of the Inria evaluation committee (CE)

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Master: E. Kaufmann, Spring 2018, Data Mining, M1 Maths/Finances, Université de Lille (36 hours)

Master: E. Kaufmann, Spring 2018, Machine Learning, M2 Maths/Finances, Université de Lille (18 hours)

Master: M. Valko, 2018/2019: Graphs in Machine Learning, 36h eqTD, M2, ENS Cachan

Master: O. Maillard, Spring 2018: Sequential Learning course, parcours DAD, 30h eqTD, Ecole Centrale Lille.

Master: O. Maillard, January 2018: Sequential Learning tutorial, Technicolor, 6h eqTD, Rennes

10.2.2. Supervision

PhD completion: Merwan Barlier, Human-in-the loop reinforcement learning for dialogue systems, started Oct. 2014, advisor: Olivier Pietquin

PhD completion: Alexandre Bérard, Deep learning for post-editing and automatic translation, started Oct. 2014, advisor: Olivier Pietquin

PhD in progress: Lilian Besson, Bandit approach to improve Internet Of Things Communications, started Oct. 2016, advisor: Émilie Kaufmann, Christophe Moy (CentraleSupélec Rennes)

PhD in progress: Ronan Fruit, Exploration-exploitation in hierarchical reinforcement learning, Inria, started Dec. 2015, advisor: Daniil Ryabko, 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: Édouard Leurent, Autonomous vehicle control: application of machine learning to contextualized path planning, started Oct. 2017, advisor: Odalric Maillard, Philippe Preux, Denis Effimov (NON-A), Wilfrid Perruquetti (NON-A)

PhD aborted: Sheikh Waqas Akhtar, Bandits for non-stationarity and structure, started Oct. 2017, advisor: Odalric Maillard, Daniil Ryabko.

PhD in progress: Pierre Perrault, Online Learning on Streaming Graphs, started Sep. 2017, advisor: Michal Valko; Vianney Perchet

PhD in progress: Mathieu Seurin, Multi-scale rewards in reinforcement learning, started Oct. 2017, advisor: Olivier Pietquin, Philippe Preux

PhD in progress: Julien Seznec, Sequential Learning for Educational Systems, started Mar. 2017, advisor: Michal Valko; Alessandro Lazaric, Jonathan Banon

PhD in progress: Xuedong Shang, Adaptive methods for optimization in stochastic environments, started Oct. 2017, advisor: Émilie Kaufmann, Michal Valko

PhD in progress: Florian Strub, Reinforcement Learning for visually grounded interaction, started Jan. 2016, advisors: Olivier Pietquin and Jeremie Mary

PhD in progress: Kiewan Villatel, Deep Learning for Conversion Rate Prediction in Online Advertising, started Oct. 2017, advisor: Philippe Preux

PhD in progress: Yannis Flet-Berliac, start Oct. 2018

PhD in progress: Hassan Saber, start Oct. 2018, Structured Multi-armed bandits, advisor: Odalric Maillard, Philippe Preux.

PhD in progress: Omar Darwiche, start Oct. 2018, Sequential Learning in Dynamic Environments, advisor: Émilie Kaufmann, Michal Valko

10.2.3. Juries

PhD and HDR juries:

- É. Kaufmann:
  - Stefan Magureanu, KTH Stockholm, February 20th, 2018
  - Valentin Reis, LIG, Grenoble, September 28th, 2018
  - Maryam Aziz, Northestearn University (Boston), December 6th, 2018
- O. Maillard: Stefan Magureanu, February 20th, 2018
- Ph. Preux:

    - Saeed Varasteh Yazdi, LIG, Grenoble
    - Fabien Vilar, Marseille
    - Merwan Barlier, Lille
- M. Valko:
    - *Pierre Ménard*, Université Toulouse 3 Paul Sabatier, June 2018, Sur la notion d'optimalité dans les problèmes de bandits stochastiques. *Reviewer*
    - *Mariana Vargas Vieyra*, Université Lille, September 2017, Adaptive graph learning with application to natural language processing. *Ph.D. mid-term evaluation reviewer*

10.3. Popularization

10.3.1. Internal or external Inria responsibilities

Philippe Preux chaired the Inria evaluation seminar of theme "Optimization, machine learning and statistical methods" in March 2018.

10.3.2. Articles and contents

- Ph. Preux interviewed for various journals ("Les échos", ...).
- Adobe research highlights M. Valko's work on online influence maximization presented (January 2018)
- Daniele Calandriello (supervised by A. Lazaric and M. Valko) wins the prize for the Best AI Thesis in France in 2018. Articles in:
    - La Voix du Nord
    - CNRS journal
    - Newstank
    - Lille1
    - Actu

10.3.3. Education

- Ph. Preux presented and animated 3 sessions on AI at the "congrès annuel du réseau national professionnel des cultures scientifique technique et industrielle" (Amcsti)

10.3.4. Interventions

- Philippe Preux:
    - presented AI related to health industry at the yearly general assembly of Eurasanté
    - presented and animated 3 sessions on AI at the "congrès annuel du réseau national professionnel des cultures scientifique technique et industrielle" (Amcsti)
    - participated to a panel at Conext forum (Lille)
- the work on Guesswhat?!:
    - has been invited to be presented on the Inria booth during The Web Conf in Lyon
    - is presented at the Inria showroom inaugurated in Dec. 2018 in Lille

10.3.5. Creation of media or tools for science outreach

- Ph. Preux was interviewed for a video about robots and AI

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] H. DE VRIES, F. STRUB, J. MARY, H. LAROCHELLE, O. PIETQUIN, A. COURVILLE.*Modulating early visual processing by language*, in "Conference on Neural Information Processing Systems", Long Beach, United States, December 2017, https://hal.inria.fr/hal-01648683

[4] 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

[5] 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

[6] 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

[7] 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

[8] 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

[9] R. MUNOS.*From Bandits to Monte-Carlo Tree Search: The Optimistic Principle Applied to Optimization and Planning*, in "Foundations and Trends in Machine Learning", 2014, vol. 7, n^{o} 1, p. 1-129, http://dx.doi.org/10.1561/2200000038

[10] 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

Publications of the year

Doctoral Dissertations and Habilitation Theses

[11] R. WARLOP.*Novel Learning and Exploration-Exploitation Methods for Effective Recommender Systems*, Lille1, October 2018, https://hal.inria.fr/tel-01915499

Articles in International Peer-Reviewed Journal

[12] B. DANGLOT, P. PREUX, B. BAUDRY, M. MONPERRUS.*Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation*, in "Empirical Software Engineering", August 2018, vol. 23, n^{o} 4, p. 2086–2119, https://arxiv.org/abs/1611.09187 [*DOI :* 10.1007/S10664-017-9571-8], https://hal.archives-ouvertes.fr/hal-01378523

[13] V. DUMOULIN, E. PEREZ, H. VRIES, F. STRUB, N. SCHUCHER, A. COURVILLE, Y. BENGIO.*Feature-wise transformations: A simple and surprisingly effective family of conditioning mechanisms*, in "Distill", July 2018, vol. 3, n^{o} 7 [*DOI :* 10.23915/DISTILL.00011], https://hal.inria.fr/hal-01841985

[14] A. DURAND, O.-A. MAILLARD, J. PINEAU.*Streaming kernel regression with provably adaptive mean, variance, and regularization*, in "Journal of Machine Learning Research", 2018, vol. 1, p. 1 - 48, https://arxiv.org/abs/1708.00768 , https://hal.archives-ouvertes.fr/hal-01927007

[15] E. KAUFMANN, T. BONALD, M. LELARGE.*A spectral algorithm with additive clustering for the recovery of overlapping communities in networks*, in "Theoretical Computer Science", September 2018, vol. 742, p. 3-26, https://hal.archives-ouvertes.fr/hal-01963868

[16] O.-A. MAILLARD.*Boundary Crossing Probabilities for General Exponential Families*, in "Mathematical Methods of Statistics", 2018, vol. 27, https://hal.archives-ouvertes.fr/hal-01737150

[17] M. S. TALEBI, O.-A. MAILLARD.*Variance-Aware Regret Bounds for Undiscounted Reinforcement Learning in MDPs*, in "Journal of Machine Learning Research", April 2018, p. 1-36, https://hal.archives-ouvertes.fr/hal-01737142

International Conferences with Proceedings

[18] Y. ABBASI-YADKORI, P. BARTLETT, V. GABILLON, A. MALEK, M. VALKO.*Best of both worlds: Stochastic & adversarial best-arm identification*, in "Conference on Learning Theory", Stockholm, Sweden, 2018, https://hal.inria.fr/hal-01808948

[19] M. AZIZ, J. ANDERTON, E. KAUFMANN, J. ASLAM.*Pure Exploration in Infinitely-Armed Bandit Models with Fixed-Confidence*, in "ALT 2018 - Algorithmic Learning Theory", Lanzarote, Spain, JMLR Workshop and Conference Proceedings, April 2018, https://arxiv.org/abs/1803.04665 , https://hal.archives-ouvertes.fr/hal-01729969

[20] M. BARLIER, R. LAROCHE, O. PIETQUIN.*Training Dialogue Systems With Human Advice*, in "AAMAS 2018 - the 17th International Conference on Autonomous Agents and Multiagent Systems", Stockholm, Sweden, International Foundation for Autonomous Agents and MultiAgent Systems (IFAAMAS), July 2018, 9, https://hal.archives-ouvertes.fr/hal-01945831

[21] P. BARTLETT, V. GABILLON, M. VALKO.*A simple parameter-free and adaptive approach to optimization under a minimal local smoothness assumption*, in "Algorithmic Learning Theory", Chicago, United States, 2019, https://hal.inria.fr/hal-01885368

[22] L. BESSON, E. KAUFMANN.*Multi-Player Bandits Revisited*, in "Algorithmic Learning Theory", Lanzarote, Spain, Mehryar Mohri and Karthik Sridharan, April 2018, https://arxiv.org/abs/1711.02317 , https://hal.inria.fr/hal-01629733

[23] L. BESSON, E. KAUFMANN, C. MOY.*Aggregation of Multi-Armed Bandits Learning Algorithms for Opportunistic Spectrum Access*, in "IEEE WCNC - IEEE Wireless Communications and Networking Conference", Barcelona, Spain, April 2018 [*DOI :* 10.1109/WCNC.2018.8377070], https://hal.inria.fr/hal-01705292

[24] A. BÉRARD, L. BESACIER, A. C. KOCABIYIKOGLU, O. PIETQUIN.*End-to-End Automatic Speech Translation of Audiobooks*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Alberta, Canada, April 2018, https://hal.archives-ouvertes.fr/hal-01709586

[25] D. Calandriello, I. Koutis, A. Lazaric, M. Valko.*Improved large-scale graph learning through ridge spectral sparsification*, in "International Conference on Machine Learning", Stockholm, Sweden, ICML 2018 - Thirty-fifth International Conference on Machine Learning, July 2018, https://hal.inria.fr/hal-01810980

[26] N. Carrara, R. Laroche, J.-L. Bouraoui, T. Urvoy, O. Pietquin.*A Fitted-Q Algorithm for Budgeted MDPs*, in "EWRL 2018 - 14th European workshop on Reinforcement Learning", Lille, France, October 2018, https://hal.archives-ouvertes.fr/hal-01928092

[27] N. Carrara, R. Laroche, J.-L. Bouraoui, T. Urvoy, O. Pietquin.*Safe transfer learning for dialogue applications*, in "SLSP 2018 - 6th International Conference on Statistical Language and Speech Processing", Mons, Belgium, October 2018, https://hal.archives-ouvertes.fr/hal-01928102

[28] R. Fruit, M. Pirotta, A. Lazaric.*Near Optimal Exploration-Exploitation in Non-Communicating Markov Decision Processes*, in "32nd Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, https://hal.inria.fr/hal-01941220

[29] R. Fruit, M. Pirotta, A. Lazaric, R. Ortner.*Efficient Bias-Span-Constrained Exploration-Exploitation in Reinforcement Learning*, in "ICML 2018 - The 35th International Conference on Machine Learning", Stockholm, Sweden, Proceedings of Machine Learning Research, July 2018, vol. 80, p. 1578-1586, https://hal.inria.fr/hal-01941206

[30] P. Gajane, T. Urvoy, E. Kaufmann.*Corrupt Bandits for Preserving Local Privacy*, in "ALT 2018 - Algorithmic Learning Theory", Lanzarote, Spain, Proceedings of Machine Learning Research, April 2018, https://hal.archives-ouvertes.fr/hal-01757297

[31] J.-B. Grill, M. Valko, R. Munos.*Optimistic optimization of a Brownian*, in "NeurIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, https://hal.inria.fr/hal-01906601

[32] J.-H. Jacobsen, A. Smeulders, E. Oyallon.*i-RevNet: Deep Invertible Networks*, in "ICLR 2018 - International Conference on Learning Representations", Vancouver, Canada, April 2018, https://arxiv.org/abs/1802.07088 , https://hal.archives-ouvertes.fr/hal-01712808

[33] E. Oyallon, E. Belilovsky, S. Zagoruyko, M. Valko.*Compressing the Input for CNNs with the First-Order Scattering Transform*, in "European Conference on Computer Vision", Munich, Germany, 2018, https://hal.inria.fr/hal-01850921

[34] M. Papini, D. Binaghi, G. Canonaco, M. Pirotta, M. Restelli.*Stochastic Variance-Reduced Policy Gradient*, in "ICML 2018 - 35th International Conference on Machine Learning", Stockholm, Sweden, Proceedings of Machine Learning Research, July 2018, vol. 80, p. 4026-4035, https://hal.inria.fr/hal-01940394

[35] E. Perez, F. Strub, H. De Vries, V. Dumoulin, A. Courville.*FiLM: Visual Reasoning with a General Conditioning Layer*, in "AAAI Conference on Artificial Intelligence", New Orleans, United States, February 2018, https://arxiv.org/abs/1707.03017 , https://hal.inria.fr/hal-01648685

[36] J. Pérolat, B. Piot, O. Pietquin.*Actor-Critic Fictitious Play in Simultaneous Move Multistage Games*, in "AISTATS 2018 - 21st International Conference on Artificial Intelligence and Statistics", Playa Blanca, Lanzarote, Canary Islands, Spain, April 2018, https://hal.inria.fr/hal-01724227

[37] J. SEZNEC, A. LOCATELLI, A. CARPENTIER, A. LAZARIC, M. VALKO.*Rotting bandits are no harder than stochastic ones*, in "International Conference on Artificial Intelligence and Statistics", Okinawa, Japan, 2019, https://hal.inria.fr/hal-01936894

[38] A. TIRINZONI, A. SESSA, M. PIROTTA, M. RESTELLI.*Importance Weighted Transfer of Samples in Reinforcement Learning*, in "ICML 2018 - The 35th International Conference on Machine Learning", Stockholm, Sweden, Proceedings of Machine Learning Research, July 2018, vol. 80, p. 4936-4945, https://hal.inria.fr/hal-01941213

Conferences without Proceedings

[39] E. KAUFMANN, W. KOOLEN, A. GARIVIER.*Sequential Test for the Lowest Mean: From Thompson to Murphy Sampling*, in "Advances in Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2018, https://arxiv.org/abs/1806.00973 , https://hal.archives-ouvertes.fr/hal-01804581

[40] E. LEURENT, Y. BLANCO, D. EFIMOV, O.-A. MAILLARD.*Approximate Robust Control of Uncertain Dynamical Systems*, in "32nd Conference on Neural Information Processing Systems (NeurIPS 2018) Workshop", Montréal, Canada, December 2018, https://hal.archives-ouvertes.fr/hal-01931744

[41] X. SHANG, E. KAUFMANN, M. VALKO.*Adaptive black-box optimization got easier: HCT only needs local smoothness*, in "European Workshop on Reinforcement Learning", Lille, France, October 2018, https://hal.inria.fr/hal-01874637

[42] F. STRUB, M. SEURIN, E. PEREZ, H. DE VRIES, J. MARY, P. PREUX, A. COURVILLE, O. PIETQUIN.*Visual Reasoning with Multi-hop Feature Modulation*, in "ECCV 2018 - 15th European Conference on Computer Vision", Munich, Germany, V. FERRARI, M. HEBERT, C. SMINCHISESCU, Y. WEISS (editors), Part of the Lecture Notes in Computer Science book series - LNCS, September 2018, vol. 11205-11220, n^{o} 11209, p. 808-831, https://arxiv.org/abs/1808.04446 , https://hal.archives-ouvertes.fr/hal-01927811

[43] R. WARLOP, A. LAZARIC, J. MARY.*Fighting Boredom in Recommender Systems with Linear Reinforcement Learning*, in "Neural Information Processing Systems", Montreal, Canada, December 2018, https://hal.inria.fr/hal-01915468

Other Publications

[44] R. ALAMI, O.-A. MAILLARD, R. FÉRAUD.*Memory Bandits: Towards the Switching Bandit Problem Best Resolution*, August 2018, MLSS 2018 - Machine Learning Summer School, Poster, https://hal.archives-ouvertes.fr/hal-01879251

[45] L. BESSON.*A Note on the Ei Function and a Useful Sum-Inequality*, July 2018, https://hal.inria.fr/hal-01847480

[46] L. BESSON.*SMPyBandits: an Experimental Framework for Single and Multi-Players Multi-Arms Bandits Algorithms in Python*, July 2018, working paper or preprint, https://hal.inria.fr/hal-01840022

[47] L. BESSON, E. KAUFMANN.*What Doubling Tricks Can and Can't Do for Multi-Armed Bandits*, February 2018, https://arxiv.org/abs/1803.06971 - working paper or preprint, https://hal.inria.fr/hal-01736357

[48] N. CARRARA, R. LAROCHE, J.-L. BOURAOUI, T. URVOY, O. PIETQUIN.*A Fitted-Q Algorithm for Budgeted MDPs*, August 2018, Workshop on Safety, Risk and Uncertainty in Reinforcement Learning. https://sites.google.com/view/rl-uai2018/, https://hal.archives-ouvertes.fr/hal-01867353

[49] G. GAUTIER, R. BARDENET, M. VALKO.*DPPy: Sampling Determinantal Point Processes with Python*, September 2018, working paper or preprint, https://hal.inria.fr/hal-01879424

[50] E. KAUFMANN, W. M. KOOLEN.*Mixture Martingales Revisited with Applications to Sequential Tests and Confidence Intervals*, October 2018, https://arxiv.org/abs/1811.11419 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01886612

[51] E. LEURENT.*A Survey of State-Action Representations for Autonomous Driving*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01908175

[52] Y. LIU, G. RADANOVIC, C. DIMITRAKAKIS, D. MANDAL, D. C. PARKES.*Calibrated Fairness in Bandits*, December 2018, working paper or preprint [*DOI :* 10.1145/NNNNNNN.NNNNNNN], https://hal.inria.fr/hal-01953314

[53] O.-A. MAILLARD, M. ASADI.*Upper Confidence Reinforcement Learning exploiting state-action equivalence*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01945034

[54] K. VILLATEL, E. SMIRNOVA, J. MARY, P. PREUX.*Recurrent Neural Networks for Long and Short-Term Sequential Recommendation*, July 2018, https://arxiv.org/abs/1807.09142 - working paper or preprint, https://hal.inria.fr/hal-01847127

[55] H. VAN HASSELT, Y. DORON, F. STRUB, M. HESSEL, N. SONNERAT, J. MODAYIL.*Deep Reinforcement Learning and the Deadly Triad*, December 2018, https://arxiv.org/abs/1812.02648 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01949304

References in notes

[56] 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

[57] R. BELLMAN.*Dynamic Programming*, Princeton University Press, 1957

[58] D. BERTSEKAS, S. SHREVE.*Stochastic Optimal Control (The Discrete Time Case)*, Academic Press, New York, 1978

[59] D. BERTSEKAS, J. TSITSIKLIS.*Neuro-Dynamic Programming*, Athena Scientific, 1996

[60] M. PUTERMAN.*Markov Decision Processes: Discrete Stochastic Dynamic Programming*, John Wiley and Sons, 1994

[61] H. ROBBINS.*Some aspects of the sequential design of experiments*, in "Bull. Amer. Math. Soc.", 1952, vol. 55, p. 527–535

[62] R. SUTTON, A. BARTO.*Reinforcement learning: an introduction*, MIT Press, 1998

[63] 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

1. Team, Visitors, External Collaborators **405**
2. Overall Objectives **407**
2.1. Introduction 407
2.2. Scientific Foundations 407
3. Research Program **407**
3.1. Introduction 407
3.2. Objective #1: Self-healing - Mining software artifacts to automatically evolve systems 407
3.2.1. Learning from software history how to design software and fix bugs 408
3.2.2. Run-time self-healing 409
3.3. Objective #2: Self-optimization - Sharing runtime behaviors to continuously adapt software 409
3.3.1. Monitoring software in the wild 410
3.3.2. Collaborative decision-making approaches 411
3.3.3. Smart reconfigurations in the large 411
4. Application Domains **412**
4.1. Introduction 412
4.2. Distributed software services for the retail industry 412
4.3. Distributed software services for the digital home 412
5. Highlights of the Year **412**
6. New Software and Platforms **413**
6.1. APISENSE 413
6.2. PowerAPI 413
6.3. Saloon 414
6.4. SPOON 414
7. New Results **415**
7.1. Software Product Lines for Setup and Adaptation of Multi-Cloud Computing Systems 415
7.2. Automated Software Repair with Patch Generation in Production 415
7.3. Flexible Framework for Elasticity in Cloud Computing 415
7.4. Semantic Interoperability in Multi-Cloud Computing Systems 416
8. Bilateral Contracts and Grants with Industry **416**
8.1. ip-label 416
8.2. Scalair 416
8.3. Davidson 416
8.4. Orange Labs #1 417
8.5. Orange Labs #2 417
8.6. Amaris 417
9. Partnerships and Cooperations **417**
9.1. Regional Initiatives 417
9.1.1. Région Hauts-de-France 417
9.1.1.1. CIRRUS 417
9.1.1.2. Alloy@Scale 418
9.1.2. Inria Lille - Nord Europe 418
9.1.2.1. ADT LibRepair 418
9.1.2.2. ADT FingerKit 418
9.1.2.3. ADT e-Lens 418
9.2. National Initiatives 418
9.2.1. ANR 418
9.2.1.1. ANR BottleNet 418
9.2.1.2. ANR SATAS 419
9.2.1.3. ANR Headwork 419

9.2.1.4. ANR Delta 419
9.2.2. Competitivity Clusters 419
9.2.3. Inria National Initiatives 419
9.2.4. CNRS Momentum 420
9.3. European Initiatives 420
9.3.1. FP7 & H2020 Projects 420
9.3.2. Collaborations in European Programs, Except FP7 & H2020 421
9.4. International Initiatives 421
9.4.1. Inria Associate Teams Not Involved in an Inria International Labs 421
9.4.2. Participation in Other International Programs 422
9.5. International Research Visitors 422
9.5.1. Visits of International Scientists 422
9.5.2. Visits to International Teams 422
10. Dissemination **422**
10.1. Promoting Scientific Activities 422
10.1.1. Scientific Events Organisation 422
10.1.2. Scientific Events Selection 423
10.1.3. Journal 424
10.1.3.1. Member of the Editorial Boards 424
10.1.3.2. Reviewer - Reviewing Activities 424
10.1.4. Invited Talks 424
10.1.5. Leadership within the Scientific Community 425
10.1.6. Scientific Expertise 425
10.1.7. Research Administration 425
10.2. Teaching - Supervision - Juries 425
10.2.1. Teaching 425
10.2.2. Supervision 427
10.2.3. Juries 427
10.3. Popularization 428
11. Bibliography **428**

Project-Team SPIRALS

Creation of the Team: 2014 January 01, updated into Project-Team: 2015 January 01

Keywords:

Computer Science and Digital Science:

A1.3. - Distributed Systems
A1.3.3. - Blockchain
A1.3.5. - Cloud
A1.4. - Ubiquitous Systems
A1.6. - Green Computing
A2.3.1. - Embedded systems
A2.3.2. - Cyber-physical systems
A2.4.2. - Model-checking
A2.5. - Software engineering
A2.5.2. - Component-based Design
A2.5.3. - Empirical Software Engineering
A2.5.4. - Software Maintenance & Evolution
A2.6.2. - Middleware
A3.1.3. - Distributed data
A3.1.9. - Database
A3.2.1. - Knowledge bases
A3.2.4. - Semantic Web
A7.2. - Logic in Computer Science

Other Research Topics and Application Domains:

B6.1. - Software industry
B6.4. - Internet of things
B6.5. - Information systems
B6.6. - Embedded systems
B8.5.2. - Crowd sourcing
B9.5.1. - Computer science
B9.5.6. - Data science
B9.10. - Privacy

1. Team, Visitors, External Collaborators

Research Scientists

Simon Bliudze [Inria, Researcher]
Pierre Bourhis [CNRS, Researcher]
Philippe Merle [Inria, Researcher, HDR]

Faculty Members

Lionel Seinturier [Team leader, Université de Lille, Professor, HDR]
Clément Quinton [Université de Lille, Associate Professor]
Laurence Duchien [Université de Lille, Professor, HDR]

Romain Rouvoy [Université de Lille, Professor, HDR]
Walter Rudametkin Ivey [Université de Lille, Associate Professor]

External Collaborator

Martin Monperrus [KTH (on leave from Université de Lille)]

Technical Staff

Jordan Bouchoucha [Université de Lille, from Oct 2018]
Aurélien Bourdon [Inria, until Sep 2018]
Antoine Canda [Inria, from Oct 2018]
Arthur d'Azemar [Inria, from Oct 2018]
Julien Decaudin [Inria, from Jun 2018]
Bénédicte Delcourt [Inria, from Sep 2018]
Christophe Gourdin [Inria]
Abderrahman Lahiaouni [Inria, from Dec 2018]
Zakaria Ournani [Inria, from Apr 2018 until Oct 2018]
Romain Sommerard [Inria, until Feb 2018]
Simon Urli [Inria, until Aug 2018]
Antoine Veuiller [Inria, until Mar 2018]

PhD Students

Zeinab Abou Khalil [Université de Lille]
Yahya Al-Dhuraibi [Scalair]
Mohammed Chakib Belgaid [Inria]
Sacha Brisset [Amaris, from Nov 2018]
Stéphanie Challita [Inria]
Benjamin Danglot [Inria]
Antonin Durey [Université de Lille, from Oct 2018]
Thomas Durieux [Inria, until Aug 2018]
Guillaume Fieni [Université de Lille]
Sarra Habchi [Inria]
Lakhdar Meftah [Inria]
Vikas Mishra [Inria, from Oct 2018]
Zakaria Ournani [Orange, from Nov 2018]
Gustavo Sousa [Université de Lille, until Feb 2018]
Marion Tommasi [Inria, from Oct 2018]
Antoine Vastel [Université de Lille]

Post-Doctoral Fellows

Alexandre Garnier [Inria, from Sep 2018]
Louis Jachiet [CNRS, from Oct 2018]
Hoang Gia Nguyen [Inria, from Nov 2018]
Zhongxing Yu [Inria, until Feb 2018]
Faiez Zalila [Inria]

Visiting Scientist

Fatima Belkouch Ameline [Université de Lille]

Administrative Assistants

Nathalie Bonte [Inria, until Apr 2018]
Karine Lewandowski [Inria, from Apr 2018]

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 [64], the NESSI Research Priorities for the next Framework Programme for Research and Technological Development FP8 [67], the Roadmap for Advanced Cloud Technologies under H2020 [65], and research roadmaps, such as [72], [63], [53].

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 [69].

[0] http://cordis.europa.eu/fp7/ict/ssai

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.

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** [70]. 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 [51].

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 [47] and [68]. 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” [62] 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 [52]. 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 behaviors 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 [71]. 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 [61] to provide enriched information to the decision level.

[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].

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 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 [50], 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 [44], 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 [49], but also the adoption of publish/subscribe abstractions [55] 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 [66] 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 [56], 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 built, 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 compétivité 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

5.1.1. Awards

Thomas Durieux, who defended his PhD thesis [11] in September 2018 on automated software repair, has obtained two best paper awards in 2018: at the 29th IEEE International Symposium on Software Reliability Engineering (ISSRE) [26] for his work on self-healing web applications based on HTML and JavaScript rewriting, and at the 6th Workshop on Software Visualization, Evolution and Maintenance (VEM) [38] for his contribution to the automatic detection of bug fixing patterns, which is a joint work with Fernanda Madeiral and colleagues from the Federal University of Uberlândia in Brazil.

Browser fingerprinting [48] has emerged as a technique to track users without their consent. Studying the implication of this technique on user privacy, and proposing software systems that can self-protect against such techniques, have become an important topic in the research activities of Spirals on self-adaptation. In 2018, two papers authored by Spirals members in the context of the ongoing PhD thesis of Antoine Vastel have been published in A* conferences of the domain of security: [36] at the 39th IEEE Symposium on Security and Privacy, and [35] at 27th USENIX Security Symposium.

In 2018, we published at the 40th International Conference on Software Engineering, Track Software Engineering in Practice, a paper on Repairnator [34], which this is the very first bot for automated software repair. This work has been conducted by Simon Urli and his coworkers in Spirals in the context of Inria ADT Librepair. An article on Repairnator has also been published on the Inria web site.

Stéphanie Challita won the Prix L'Oréal-UNESCO For Women in Science. She is among the 30 laureates rewarded out of 900 candidates. Stéphanie Challita has defended her PhD thesis in December 2018 on FCLOUDS, a framework to achieve semantic interoperability in multi-cloud computing systems.

Xscalibur is a transfer project led by Christophe Gourdin and Philippe Merle that aims at creating a startup company with innovative solutions for the management of multi-cloud systems. The project has been selected by Serre Numérique Valenciennes. The research activities that led to this project have been, in part, conducted in the context of the joint lab between Spirals and the Scalair cloud hosting company. Several papers in relation with these research activities have also been published in 2018: [30], [25], [24], [13], [23].

BEST PAPERS AWARDS :

[26]
T. DURIEUX, Y. HAMADI, M. MONPERRUS.*Fully Automated HTML and Javascript Rewriting for Constructing a Self-healing Web Proxy*, in "Proceedings of the 29th IEEE International Symposium on Software Reliability Engineering (ISSRE 2018)", Memphis, United States, October 2018, https://arxiv.org/abs/1803.08725 [*DOI :* 10.1109/ISSRE.2018.00012], https://hal.inria.fr/hal-01746141

[38]
F. MADEIRAL, T. DURIEUX, V. SOBREIRA, M. MAIA.*Towards an automated approach for bug fix pattern detection*, in "VEM '18 - Proceedings of the VI Workshop on Software Visualization, Evolution and Maintenance", São Carlos, Brazil, September 2018, https://arxiv.org/abs/1807.11286 , https://hal.archives-ouvertes.fr/hal-01851813

6. New Software and Platforms

6.1. APISENSE

KEYWORDS: Mobile sensing - Crowd-sensing - Mobile application - 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: Antoine Veuiller, Christophe Ribeiro, Julien Duribreux, Nicolas Haderer, Romain Rouvoy, Romain Sommerard and Lakhdar Meftah
- Partner: Université de Lille
- Contact: Romain Rouvoy
- URL: https://apisense.io

6.2. PowerAPI

KEYWORDS: Energy efficiency - 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: Adel Noureddine, Loïc Huertas, Maxime Colmant and Romain Rouvoy
- Contact: Romain Rouvoy
- URL: http://powerapi.org

6.3. 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: Clément 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

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: Gérard Paligot, Lionel Seinturier, Martin Monperrus and Nicolas Petitprez
- Contact: Martin Monperrus
- URL: http://spoon.gforge.inria.fr

7. New Results

7.1. Software Product Lines for Setup and Adaptation of Multi-Cloud Computing Systems

In 2018, in the domain of cloud computing, we proposed a new software product line-based approach for managing the variability in order to automate the setup and adaptation of multi-cloud environments. Building such systems is still very challenging and time consuming due to the heterogeneity across cloud providers' offerings and the high-variability in the configuration of cloud providers. This variability is expressed by the large number of available services and the many different ways in which they can be combined and configured. In order to ensure correct setup of a multi-cloud environment, developers must be aware of service offerings and configuration options from multiple cloud providers. Our results enable to automatically generate a configuration or reconfiguration plan for a multi-cloud environment from a description of its requirements. The conducted experiments aim to assess the impact of the approach on the automated analysis of feature models and the feasibility of the approach to automate the setup and adaptation of multi-cloud environments. These results have been obtained in the context of the PhD thesis of Gustavo Sousa [12] defended in June 2018.

7.2. Automated Software Repair with Patch Generation in Production

In 2018, in the domain of automated software repair, we proposed new patch generation techniques. Patch creation is one of the most important actions in the life cycle of an application. Creating patches is a time-consuming task. Not only because it is difficult to create a sound and valid patch, but also because it requires the intervention of humans. Our work proposes new patch generation techniques that remove the human intervention. Our idea is to put as close as possible the patch generation in the production environment. We adopt this approach because the production environment contains all the data and human interactions that lead to the bug. We show how to exploit this data to detect bugs, generate and validate patches. We evaluate this approach on seven different benchmarks of real bugs collected from open-source projects. During the evaluation, we are particularly attentive to the number of generated patches, to their correctness, readability and to the time required for generating them. Our evaluation shows the applicability and feasibility of our approach to generate patches in the production environment without the intervention of a developer. These results have been obtained in the context of the PhD thesis of Thomas Durieux [11] defended in September 2018.

7.3. Flexible Framework for Elasticity in Cloud Computing

In 2018, in the domain of cloud computing, we proposed a new framework for managing elasticity. The main factor motivating the use of cloud is its ability to provide resources according to the customer needs or what is referred to as elasticity. Adapting cloud applications during their execution according to demand variation is nevertheless a challenging task. In addition, cloud elasticity is diverse and heterogeneous because it encompasses different approaches, policies, purposes, etc. In this work, three contributions are proposed: (1) an up-to-date state-of-the-art of the cloud elasticity for both virtual machines and containers, (2) ELASTICDOCKER, an approach to manage container elasticity including vertical elasticity, live migration, and elasticity combination between different virtualization techniques, and (3) MODEMO, a new unified standard-based, model-driven, highly extensible and reconfigurable framework that supports multiple elasticity policies, vertical and horizontal elasticity, different virtualization techniques and multiple cloud providers. These results have been obtained in the context of the PhD thesis of Yahya Al-Dhuraibi defended in December 2018.

7.4. Semantic Interoperability in Multi-Cloud Computing Systems

In 2018, in the domain of cloud computing, we proposed two major results related to semantic interoperability. First, an approach based on reverse-engineering to extract knowledge from the ambiguous textual documentation of cloud APIs and to enhance its representation using MDE techniques has been proposed. This approach is applied to Google Cloud Platform (GCP), where we provide GCP Model, a precise model-driven specification for GCP. GCP Model is automatically inferred from GCP textual documentation, conforms to the OCCIWARE METAMODEL and is implemented within OCCIWARE STUDIO. It allows one to perform qualitative and quantitative analysis of the GCP documentation. Second, we have proposed the FCLOUDS framework to achieve semantic interoperability in multi-clouds, i.e., to identify the common concepts between cloud APIs and to reason over them. The FCLOUDS language is a formalization of OCCI concepts and operational semantics in Alloy formal specification language. To demonstrate the effectiveness of the FCLOUDS language, we formally specify thirteen case studies and verify their properties. Then, thanks to formal transformation rules and equivalence properties, we draw a precise alignment between my case studies, which promotes semantic interoperability in multi-clouds. These results have been obtained in the context of the PhD thesis of Stéphanie Challita defended in December 2018.

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. Scalair

Participants: Yahya Al-Dhuraibi, Philippe Merle [correspondant].

This collaboration (2015–18) aims at proposing a framework to deal with elasticity in cloud computing environments. This framework must cover all kinds 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 PhD thesis of Yahya Al-Dhuraibi defended in December 2018.

8.3. Davidson

Participants: Romain Rouvoy [correspondant], Lionel Seinturier.

This collaboration (2017–20) aims at proposing new solutions for optimizing the energy footprint of ICT software infrastructures. We want to be able to measure and assess the energy footprint of ICT systems while preserving various quality of service parameters, such as performance and security. We aim at proposing a testbed for assessing the energy footprint of various programming languages. This testbed will also incorporate frameworks for web and mobile programming. Finally, we want to be able to issue recommendations to developers in order to assist them in improving the energy footprint of their programs. This collaboration will take advantage of the POWERAPI software library.

The PhD of Mohammed Chakib Belgaid takes place in the context of this collaboration.

8.4. Orange Labs #1

Participants: Philippe Merle [correspondant], Lionel Seinturier.

This collaboration (2017–18) aims at defining a computational model for software infrastructures layered on top of virtualized and interconnected cloud resources. This computational model will provide application programming and management facilities to distributed applications and services. This computational model will define a pivot model that will enable the interoperability of various existing and future standards for cloud systems such as OCCI and TOSCA. This pivot model will be defined with the Alloy specification language [54]. This collaboration takes advantage of the expertise that we are developing since several years on reconfigurable component-based software systems [66], on cloud systems [60], and on the Alloy specification language [58].

This collaboration with Orange Labs is a joint project with Jean-Bernard Stefani from the Spades Inria project-team.

8.5. Orange Labs #2

Participants: Zakaria Ournani, Romain Rouvoy [correspondant], Lionel Seinturier.

This collaboration (2018–21) aims at proposing new solutions for modeling the energy efficiency of software systems and to design and implement new methods for measuring and reducing the energy consumption of software systems at development time. We especially target software systems deployed on cloud environments.

The CIFRE PhD of Zakaria Ournani takes place in the context of this collaboration.

8.6. Amaris

Participants: Sacha Brisset, Romain Rouvoy [correspondant], Lionel Seinturier.

This collaboration (2018–21) aims at proposing new solutions for automatically spotting and fixing recurrent user experience issues in web applications. We are interested in developing an autonomic framework that learns and classifies the behaviors and figures out causality links between data such as web GUI events, support tickets and user feedback, source version management events (e.g. recent commits). The ultimate objective is to implement an AI-powered recommendation system to guide the maintenance and even to automatically predict and solve user issues.

The CIFRE PhD of Sacha Brisset takes place in the context of this collaboration.

9. Partnerships and Cooperations

9.1. Regional Initiatives

9.1.1. Région Hauts-de-France

9.1.1.1. CIRRUS

Participants: Yahya Al-Dhuraibi, Stéphanie Challita, Guillaume Fieni, Alexandre Garnier, Christophe Gourdin, Philippe Merle [correspondant], Romain Rouvoy, Lionel Seinturier, Faiez Zalila.

CIRRUS is an 3-year (2017–20) joint team with the Scalair cloud operator and architect company funded by the Hauts-de-France region. The CIRRUS joint team is developing novel solutions in the domains of the on demand configuration of heterogeneous cloud resources, the management of cloud elasticity for all deployed services (SaaS, PaaS, IaaS) in order to guarantee quality of service and user quality of experience, and the taming of financial costs of cloud infrastructures.

9.1.1.2. Alloy@Scale

Participants: Philippe Merle [correspondant], Romain Rouvoy, Lionel Seinturier, Zakaria Ournani.

Alloy@Scale is a 12-month project funded in the context of CPER Data program. Alloy@Scale aims at overcoming the limits of the formal verification of large software systems specified with the Alloy formal specification language. For that, the program combines the Grid'5000 infrastructure and the Docker container technology.

9.1.2. Inria Lille - Nord Europe

9.1.2.1. ADT LibRepair

Participants: Benjamin Danglot, Martin Monperrus, Lionel Seinturier [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 [57] and Benoit Cornu [46].

9.1.2.2. ADT FingerKit

Participants: Antoine Canda, Walter Rudametkin Ivey [correspondant], Antoine Vastel.

ADT FingerKit (2018–20) is a technology development initiative supported by the Inria Lille - Nord Europe Center that focuses on the design and development of a new and enhanced version of the AmIUnique platform. AmIUnique is a data collection and analysis platform to better understand, analyze and vulgarize the uses and threats of browser fingerprinting. This initiative led by Inria is a key asset to better understand novel techniques that threatens the user privacy on Internet. This ADT builds on our first results with the ongoing PhD thesis of Antoine Vastel.

9.1.2.3. ADT e-Lens

Participants: Arthur d'Azémar, Guillaume Fieni, Romain Rouvoy [correspondant].

ADT e-Lens (2018–20) is a technology development initiative supported by the Inria Lille - Nord Europe Center that aims at extending the PowerAPI energy monitoring library that we develop in the team since 2011. The extension deals with the integration of new power models (for GPU, disk, network interface), the implementation of a self-optimization algorithm, the port of the platform to embedded systems running with Raspberry Pi, ROS and Android, and the implementation of an active learning algorithm for power models. This ADT builds on our results with the defended PhD theses of Adel Nouredine [59] and Maxime Colmant [45], and with the ongoing PhD thesis of Guillaume Fieni.

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 (2015–19) funded by ANR. 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 (2015–19) funded by ANR. 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 their 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.

9.2.1.3. ANR Headwork

Participants: Pierre Bourhis [correspondant], Marion Tommasi.

Headwork is a 48-month project (2016–21) funded by ANR. The main objective of Headwork 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. This project is coordinated by D. Gross-Amblard from the Druid Team (Rennes 1). Other partners include the Dahu team (Inria Saclay), Sumo (Inria Bretagne), and Links (Inria Lille) with J. Nierhen and M. Sakho.

9.2.1.4. ANR Delta

Participant: Pierre Bourhis [correspondant].

Delta is a 48-month project (2016–21) funded by ANR. The project focuses on the study of logic, transducers and automata. In particular, it aims at extending classical framework to handle input/output, quantities and data. This project is coordinated by M. Zeitoun from LaBRI. Other partners include LIF (Marseille), IRIF (Paris-Diderot), and D. Gallois from the Inria Lille Links team.

9.2.2. Competitivity Clusters

9.2.2.1. FUI StoreConnect

Participants: Aurélien Bourdon, Julien Duribreux, Romain Rouvoy, Lionel Seinturier [correspondant], Antoine Veuiller.

StoreConnect is a 36-month project (2016–19) funded by FUI and labelled by the PICOM (Pôle des Industries du COMmerce) competitivity cluster. The partners are Tevolys, Ubudu (leader), 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. Inria National Initiatives

9.2.3.1. Inria IPL BetterNet

Participants: Lakhdar Meftah, Romain Rouvoy [correspondant], Romain Sommerard, Antoine Veuiller.

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, and (3) new value-added services will be proposed to end-users. IPL BetterNet is led by Isabelle Chrisment (Inria Madynes), with the participation of the Diana, Dionysos, Inria Chile, Muse, and Spirals Inria project-teams, as well as the ARCEP French agency and the ip-label company.

9.2.4. CNRS Momentum

9.2.4.1. Manage Your Data Without Information Leakage

Participants: Pierre Bourhis [correspondant], Louis Jachiet.

“Gérer vos données sans fuite d’information” is a 3-year (2018–20) project granted in the context of the CNRS-Momentum call for projects. Data manipulated by modern applications are stored in large databases. To protect these pieces of data, security policies limit a user’s access to what she is allowed to see. However, by using the semantics of the data, a user can deduce information that she was not supposed to have access to. The goal of this project is to establish methods and tools for understanding and detecting such data leaks.

9.3. European Initiatives

9.3.1. FP7 & H2020 Projects

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].

Program: EIT Digital.

Project acronym: MCS.

Project title: Multi-Cloud Studio.

Duration: 12 months (2018).

Coordinator: Inria.

Other partners: Santer Reply (Italy) and Scalair (France).

Abstract: The goal of this EIT Digital activity is to create the Xscalibur start-up and develop its first Multi-Cloud Studio product. This product is a model-driven graphical interface to design, deploy and administrate multi-cloud systems based on Amazon Web Service, OpenStack, and VMware.

Participants: Jamal Boudjaj, Rida Darmal, Julien Decaudin, Bénédicte Delcourt, Christophe Gourdin, Philippe Merle [correspondant].

Program: EIT Digital.

Project acronym: Bankable.

Project title: Bankable, Deliver the best mobile banking services and customer experience

Duration: 12 months (2018).

Coordinator: Alfstore.

Other partners: Inria, Cefriel

Abstract: BANKABLE helps consumer banks continuously deliver the best mobile banking experience. The platform will anonymously analyze data gathered from mobile sensors, user context and interactions. Enriched by online customers' feedbacks, the bank will find deeper insights & drive real-time actions.

Participants: Romain Rouvoy [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: 2017

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. PHC Zenon Cyprus - Project RRI-MobDev

Participants: Sarra Habchi, Lakhdar Meftah, Mohammad Naseri, Romain Rouvoy [correspondant], Walter Rudametkin Ivey, Romain Sommerard, Antoine Vastel.

RRI-MobDev (*Responsible Research and Innovation for Mobile Application Development*) is a 2-year (2017–2018) bilateral collaboration with UCLan Cyprus, an overseas campus of the University of Central Lancashire. Mobile applications are part of a complex ecosystem involving various stakeholders (developers, users, app stores, etc.) exposed to various threats, including not only malware, but also potential information leaks through the continuous interactions with remote servers. This project aims to study and alleviate this problem by intervening both with the users and the developers of mobile apps, with an aim of enabling a cleaner, safer and more responsible mobile app ecosystem.

9.5. International Research Visitors

9.5.1. Visits of International Scientists

Rajkumar Buyya, Redmond Barry Distinguished Professor and Director of the Cloud Computing and Distributed Systems (CLOUDS) Laboratory at the University of Melbourne, visited us in February 2018.

9.5.1.1. Internships

Mohammad Naseri, MSc. Student in Computer Science from Saarland University, Germany, visited us for 3 months, from November 2017 to January 2018.

Chaima Chakhaba, MSc. Student in Computer Science from ESI Alger, Algeria, visited us for 9 months, from December 2017 to August 2018.

9.5.2. Visits to International Teams

9.5.2.1. Research Stays Abroad

Thomas Durieux, PhD Student, spent 3 months from April to June 2018 in KTH, Sweden.

10. Dissemination

10.1. Promoting Scientific Activities

10.1.1. Scientific Events Organisation

10.1.1.1. General Chair, Scientific Chair

Simon Bliudze

- co-chair International Conference on Verification and Evaluation of Computer and Communication Systems (VECoS)
- co-chair International Workshop on Methods and Tools for Rigorous System Design (MeTRiD)

Laurence Duchien

- IEEE TCSE Distinguished Women in Science and Engineering Leadership Award Committee
- co-chair ECSA Track for Women in Software Architecture

10.1.2. Scientific Events Selection

10.1.2.1. Member of the Conference Program Committees

Simon Bliudze

- International Conference on Formal Aspects of Component Software (FACS)
- International Symposium on Formal Approaches to Parallel and Distributed Systems (4PAD)
- Interaction and Concurrency Experience International Workshop (ICE)
- Euromicro Conference on Software Engineering and Advanced Applications (SEAA)
- International Conference on Coordination Models and Languages (COORDINATION)
- FME Workshop on Formal Methods in Software Engineering (FormaliSE)
- International Workshop on Foundations of Coordination Languages and Self-Adaptative Systems (Foclasa)

Pierre Bourhis

- International Joint Conferences on Artificial Intelligence (IJCAI)
- AAAI Conference on Artificial Intelligence

Laurence Duchien

- International Workshop on Variability Modelling of Software-Intensive Systems (VAMOS)
- Live Adaptation of Software Systems (LASSY)
- International Workshop on Software Qualities and their Dependencies (SQUADE)
- Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems (WDES)
- Doctoral Symposium ECSA
- ICSE Poster Session
- International Workshop on Software Engineering Research and Industrial Practise (ICSE/SER&IP)
- Conference on Software Engineering Education and Training (CSEE&T)
- European Conference on Software Architecture (ECSA)

Philippe Merle

- International Conference on Cooperative Information Systems (CoopIS)
- International Conference on Cloud Computing Work-in-Progress Track (IEEE CLOUD)
- International Conference on Web Engineering (ICWE)
- International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM)
- International Conference on Cloud Computing, GRIDs, and Virtualization (CLOUD COMPUTING)
- International Conference on Advanced Service Computing (SERVICE COMPUTATION)
- Workshop on CrossCloud Infrastructures & Platforms (CrossCloud)

Clément Quinton

- International Workshop on Software Engineering for Systems-of-Systems (SESOS)
- Euromicro Conference on Software Engineering and Advanced Applications (SEAA)
- International Systems and Software Product Line Conference (SPLC)

Romain Rouvoy

- IEEE/ACM International Conference on Automated Software Engineering (ASE), demonstration track
- ACM/USENIX European Conference on Computer Systems (Eurosys)
- IEEE International Conference on Cloud Engineering (IC2E)
- IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft)
- ACM Symposium on Applied Computing (SAC), DADS track

Lionel Seinturier

- International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS)
- IEEE/ACM International Conference on Mobile Software Engineering and Systems (MOBILESoft)
- ACM Symposium on Applied Computing (SAC) SA-TTA & OS Tracks
- Australasian Software Engineering Conference (ASWEC)
- International Conference on Software Technologies (ICSOFT)
- International Conference on Service Oriented Computing (ICSOC)
- ACM International Conference on Management of Emergent Digital EcoSystems (MEDES)

Faiez Zalila

- International Conference on Verification and Evaluation of Computer and Communication Systems (VECoS)

10.1.3. Journal

10.1.3.1. Member of the Editorial Boards

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

Yahya Al-Dhuraibi: IEEE Transactions on Services Computing (TSC), The Computer Journal of Oxford University Press.

Simon Bliudze: Elsevier Science of Computer Programming (SCICO).

Stéphanie Challita: Wiley Software: Practice & Experience (SPE), Wiley Concurrency and Computation: Practice and Experience (CCPE).

Laurence Duchien: Elsevier Journal of Systems and Software (JSS), IEEE Transaction on Cloud Computing (TCC).

Philippe Merle: Elsevier Future Generation Computer Systems (FGCS), Wiley Software: Practice and Experience (SPE), Elsevier Sustainable Computing, Informatics and Systems (SUSCOM), Wiley Concurrency and Computation: Practice and Experience (CCPE), IGI Global International Journal of Web Services Research (IJWSR).

Romain Rouvoy: Elsevier Information and Software Technology (IST).

Lionel Seinturier: Springer Software Quality Journal (SQJ), Elsevier Sustainable Computing, Informatics and Systems (SUSCOM), Springer Empirical Software Engineering (EMSE), Elsevier Future Generation Computer Systems (FGCS).

Faiez Zalila: Elsevier Future Generation Computer Systems (FGCS), IEEE ACCESS, Springer Journal of Internet Services and Applications (JISA).

10.1.4. Invited Talks

Philippe Merle: International Conference on Cloud and Robotics (ICCR), invited talk.

Faiez Zalila: Cloudification of the Internet of Things Conference (CIoT), tutorial.

10.1.5. Leadership within the Scientific Community

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 Professor in Computer Science at Université du Littoral Côte d’Opale. She was scientific expert for DGRI PHC Ulysse, NSERC-Canada, PHC Van Gogh Netherlands, PHC Amadeus Austria, Programme Emergence Ville de Paris Vinci, Prix de thèse GDR GPL, Prix de valorisation de la MESHS.

Philippe Merle was member of the recruitment committee for junior researcher (CRCN) at Inria Lille - Nord Europe, and Associate Professor at Télécom SudParis. He was scientific reviewer for ANR.

Romain Rouvoy was member of the recruitment committees for associate professor in Computer Science at ENSEIRB MATMECA (Bordeaux), Université de Nantes and Université de Rennes. He was also a member of the scientific evaluation committee of ANR (CES25). He was scientific expert for DRRT IDF, ADEME PERFECTO and FET-OPEN - Novel Ideas for Radically New Technologies - *Research and Innovation Actions* (RIA).

Lionel Seinturier was member of the recruitment committee for Professor in Computer Science at CNAM Paris, and Associate Professor at University of Lille. He was member of the Promotion Committee to Full Professor for Charles University, Prague. He was scientific expert for ANRT, ECOS Nord, DRRT IDF, PICOM/Silab Retail Booster.

10.1.7. Research Administration

Laurence Duchien is member of the CNRS CoCNRS section 6 committee, and of the “bureau” of this committee. She is elected member-at-large for the IEEE Technical Council of Software Engineering (TCSE) 2016–18.

Philippe Merle is member of the Inria scientific board, president of the CUMI (Comité des Utilisateurs des Moyens Informatiques), permanent secretary of the CLHSCT (Comité Local d’Hygiène, de Sécurité et de Conditions de Travail), and member of the centre committee for the Inria Lille - Nord Europe research center. He is member of the steering committee of the Inria’s continuous integration service. He is member of the steering committee of CIEL (Conférence en Ingénierie du Logiciel).

Romain Rouvoy is an elected member of the “bureau” of the ACM SIGOPS French chapter (ASF) and elected member of the administrative council of SpecifCampus.

Lionel Seinturier is president of the CDT (Commission Développement Technologique), and member of the BCEP (Bureau du Comité des Équipes-Projets), for the Inria Lille - Nord Europe research center. He is Scientific Advisor for the evaluation of ICT research laboratories at the Hcéres.

10.2. Teaching - Supervision - Juries

10.2.1. Teaching

Simon Bliudze is, in addition to his tenure junior research position at Inria, chargé d’enseignement at École Polytechnique, Palaiseau, France, in the Department of Computer Sciences (DIX).

- INF411: Les bases de la programmation et de l’algorithmique, 40h, 2e année du Cycle ingénieur
- INF442: Traitement des données massives, 40h, 2e année du Cycle ingénieur

Pierre Bourhis is, in addition to his tenure junior research position at CNRS, chargé d'enseignement spécialité Sciences des données at École Polytechnique, Palaiseau, France, in the Department of Computer Sciences (DIX).

- École Polytechnique, Bases de données, 18h, Cycle Polytechnique Info553
- École Polytechnique, Modal Graphe Géant, 36h
- École Centrale de Lille, Master DAD, Bases de données, 12h

Laurence Duchien teaches at the University of Lille in the FST faculty. She headed the Carrières et Emplois service. She was Director of Doctoral Studies for Computer Science in the Doctoral School Engineering Science (SPI) - ComUE Lille Nord de France. She is now project leader for the doctoral studies at University of Lille.

- Software Project Management, 50h, Level M2, Master MIAGE
- Design of distributed applications, 42h, Level M1, 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

Clément Quinton teaches at the University of Lille in the FST faculty.

- Introduction to Computer Science, 46.5h, Level L1, Licence of Computer Science
- Object-oriented programming, 36h, Level L2, Licence of Computer Science
- Object-oriented design, 42h, Level L3, Licence of Computer Science
- Design of distributed applications, 42h, Level M1, Master of Computer Science
- Advanced design of distributed applications, 37.5h, Level M2, Master MIAGE
- Infrastructure and frameworks for the Internet, 33.75h, Level M2, Master of Computer Science
- Software product lines, 7.5h, Level M2, Master of Computer Science
- Suivi de stages et de projets, 30h, Licence and Master of Computer Science

Romain Rouvoy teaches at the University of Lille in the FST faculty. He heads the Master of Computer Science program at the University of Lille.

- Design of distributed applications, 12h, Level M1, Master of Computer Science
- Object-oriented design, 4h, Level L3, Licence of Computer Science
- Suivi de projets, 20h, Level M2, Master of Computer Science

Walter Rudametkin Ivey teaches at the University of Lille in the Polytech engineering school.

- GIS4 Programmation par Objets, 32h
- GIS4 Architectures Logicielles, 26h
- GIS2A3 (apprentissage) Projet programmation par Objet, 24h
- IMA2A4 (apprentissage) Conception Modélisation Objet, 24h
- IMA3 Programmation Avancée, 62h
- GBIAAL4 Bases de données, 22h
- GIS5 Suivi de projets, 42h
- GIS2A (apprentissage) Suivi d'apprentis, 28h

Lionel Seinturier teaches at the University of Lille in the FST faculty. He heads the Computer Science Department at the Faculty of Science and Technology of the University of Lille.

- Conception d'Applications Réparties, 50h, Level M1, Master MIAGE
- Infrastructures et Frameworks Internet, 70h, Level M2 E-Services IAGL TIIR, Master of Computer Science

10.2.2. Supervision

- PhD: Yahya Al-Dhuraibi, Flexible Framework for Elasticity in Cloud Computing, University of Lille, December 2018, Philippe Merle.
- PhD: Stéphanie Challita, Inferring Models from Cloud APIs and Reasoning over Them: a Tooled and Formal Approach, University of Lille, December 2018, Philippe Merle.
- PhD: Thomas Durieux, From Runtime Failures to Patches: Study of Patch Generation in Production, University of Lille, September 2018, Lionel Seinturier & Martin Monperrus. [11]
- PhD: Gustavo Sousa, A Software Product Lines-Based Approach for the Setup and Adaptation of Multi-Cloud Environments, University of Lille, June 2018, Laurence Duchien & Walter Rudametkin Ivey. [12]
- PhD in progress: Mohammed Chakib Belgaid, Développement durable des logiciels vers une optimisation énergétique de bout en bout des systèmes logiciels, January 2018, Romain Rouvoy & Lionel Seinturier.
- PhD in progress: Vikas Mishra, Collaborative Strategies to Protect Against Browser Fingerprinting, October 2018, Romain Rouvoy & Walter Rudametkin & Lionel Seinturier.
- PhD in progress: Marion Tommasi, Collaborative Data-centric Workflows: Towards Knowledge Centric Workflows and Integrating Uncertain Data, October 2018, Pierre Bourhis & Lionel Seinturier.
- PhD in progress: Antonin Durey, Leveraging Browser Fingerprinting to Fitht Fraud on the Web, October 2018, Romain Rouvoy & Walter Rudametkin & Lionel Seinturier.
- PhD in progress: Sacha Brisset, Automatic Spotting and fixing or Recurrent user Experience issues. Detecting and Fixing Anomalies by applying Machine Learning on user Experience Data, November 2018, Lionel Seinturier & Romain Rouvoy & Renaud Pawlak & Yoann Couillec.
- PhD in progress: Zakaria Ournani, Eco-conception des logiciels : modélisation de l'efficience énergétique des logiciels et conception d'outils pour mesurer et réduire leur consommation d'énergie, November 2018, Romain Rouvoy.
- PhD in progress: Zeinab Abou Khalil, November 2017, Laurence Duchien & Clément Quinton, co-supervision with Tom Mens (University of Mons, Belgium).
- PhD in progress: Guillaume Fieni, GreenData : Vers un traitement efficient et éco-responsable des grandes masses de données numériques, October 2017, Romain Rouvoy & Lionel Seinturier.
- PhD in progress: Benjamin Danglot, Software Testing Amplification, December 2016, Martin Monperrus & Lionel Seinturier.
- PhD in progress: Lakhdar Meftah, Cartography of the Quality of Experience for Mobile Internet Access, November 2016, Romain Rouvoy, co-supervision with Isabelle Chrisment (Inria Madynes).
- PhD in progress: Sarra Habchi, Une supervision de contexte sensible à la confidentialité pour les développements logiciels en crowdsource, October 2016, Romain Rouvoy.
- PhD in progress: Antoine Vastel, Cartographie de la qualité d'expérience pour l'accès à l'internet mobile, October 2016, Romain Rouvoy & Walter Rudametkin.

10.2.3. Juries

Simon Bliudze

- Maaz Mashood Mohiuddin (EPFL, Switzerland), examiner
- Emmanouela Stachtiari (Aristhotle U. of Thessaloniki, Greece), examiner

Laurence Duchien

- HDR Chouki Tibermacine (U. Montpellier), reviewer
- HDR Thomas Ledoux (U. Nantes), reviewer
- HDR Nikolaos Georgantas (U. Pierre et Marie Curie), examiner
- HDR Matthieu Roy (U. Toulouse), president
- HDR Raul Mazo (U. Paris 1 Sorbonne), reviewer
- Anne-Lucie Vion (U. Grenoble), reviewer
- Van Cam Pham, CEA, examiner
- Benjamin Allaert (U. Lille), president
- Farzan Kalantari (U. Lille), president
- Eddy Habach (U. Nice Côte d'Azur), examiner
- Mathieu Allon (U. Lille), president
- Juan Carlos Munoz (U. Paris 1 Sorbonne)
- Paola Andreas Gomez Barreto (U. Grenoble), examiner

Philippe Merle

- HDR Sylvain Vauttier (U. Montpellier), reviewer
- Nabila Belhaj (Télécom SudParis), examiner

Lionel Seinturier

- Mohamed Mosli (Telecom SudParis), president
- William Excoffon (U. Toulouse), president
- Alex Palesandro (U. Lyon 3), president
- Bassirou Ngom (Sorbonne U.), reviewer
- Gabriel Campeanu, (Märladalen U., Sweden), examiner
- Isabelle Astic (CNAM Paris), examiner

10.3. Popularization

10.3.1. Interventions

Lionel Seinturier participated in October to the Chercheur itinérant event organized by the Inria Lille Nord Europe research center in the context of La fête de la science. Two classrooms (2nde and terminale) have been visited. The theme of the visit was crowdsensing and data gathering from mobile devices.

11. Bibliography

Major publications by the team in recent years

[1] S. Abiteboul, P. Bourhis, V. Vianu.*Explanations and Transparency in Collaborative Workflows*, in "PODS 2018 - 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles Of Database Systems", Houston, Texas, United States, June 2018, https://hal.inria.fr/hal-01744978

[2] Y. Al-Dhuraibi, F. Paraiso, N. Djarallah, P. Merle.*Elasticity in Cloud Computing: State of the Art and Research Challenges*, in "IEEE Transactions on Services Computing (TSC)", March 2018, vol. 11, n^{o} 2, p. 430-447 [*DOI :* 10.1109/TSC.2017.2711009], https://hal.inria.fr/hal-01529654

[3] 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

[4] T. Durieux, Y. Hamadi, M. Monperrus.*Fully Automated HTML and Javascript Rewriting for Constructing a Self-healing Web Proxy*, in "Proceedings of the 29th IEEE International Symposium on Software Reliability Engineering (ISSRE 2018)", Memphis, United States, October 2018, https://arxiv.org/abs/1803.08725 , https://hal.inria.fr/hal-01746141

[5] S. Habchi, X. Blanc, R. Rouvoy.*On Adopting Linters to Deal with Performance Concerns in Android Apps*, in "ASE18 - Proceedings of the 33rd IEEE/ACM International Conference on Automated Software Engineering", Montpellier, France, ACM Press, September 2018, vol. 11 [*DOI :* 10.1145/3238147.3238197], https://hal.inria.fr/hal-01829135

[6] 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

[7] E. Stachtiari, A. Mavridou, P. Katsaros, S. Bliudze, J. Sifakis.*Early validation of system requirements and design through correctness-by-construction*, in "Journal of Systems and Software", 2018, vol. 145, p. 52-78 [*DOI :* 10.1016/j.jss.2018.07.053], https://hal.archives-ouvertes.fr/hal-01873999

[8] S. Urli, Z. Yu, L. Seinturier, M. Monperrus.*How to Design a Program Repair Bot? Insights from the Repairnator Project*, in "40th International Conference on Software Engineering, Track Software Engineering in Practice", Gothenburg, Sweden, May 2018, p. 95-104, update for oadoi on Nov 02 2018 [*DOI :* 10.1145/3183519.3183540], https://hal.inria.fr/hal-01691496

[9] A. Vastel, P. Laperdrix, W. Rudametkin, R. Rouvoy.*FP-STALKER: Tracking Browser Fingerprint Evolutions*, in "IEEE S&P 2018 - 39th IEEE Symposium on Security and Privacy", San Francisco, United States, B. Parno, C. Kruegel (editors), Proceedings of the 39th IEEE Symposium on Security and Privacy (S&P), IEEE, May 2018, p. 728-741 [*DOI :* 10.1109/SP.2018.00008], https://hal.inria.fr/hal-01652021

[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] T. Durieux.*From Runtime Failures to Patches: Study of Patch Generation in Production*, University of Lille, September 2018, https://tel.archives-ouvertes.fr/tel-01883676

[12] G. Sousa.*A Software Product Lines-Based Approach for the Setup and Adaptation of Multi-Cloud Environments*, Université de Lille, June 2018, https://hal.inria.fr/tel-01808779

Articles in International Peer-Reviewed Journal

[13] Y. AL-DHURAIBI, F. PARAISO, N. DJARALLAH, P. MERLE. *Elasticity in Cloud Computing: State of the Art and Research Challenges*, in "IEEE Transactions on Services Computing (TSC)", March 2018, vol. 11, n^o 2, p. 430-447 [*DOI :* 10.1109/TSC.2017.2711009], https://hal.inria.fr/hal-01529654

[14] M. COLMANT, R. ROUVOY, M. KURPICZ, A. SOBE, P. FELBER, L. SEINTURIER. *The Next 700 CPU Power Models*, in "Journal of Systems and Software", July 2018, https://hal.inria.fr/hal-01827132

[15] B. DANGLOT, P. PREUX, B. BAUDRY, M. MONPERRUS. *Correctness Attraction: A Study of Stability of Software Behavior Under Runtime Perturbation*, in "Empirical Software Engineering", August 2018, vol. 23, n^o 4, p. 2086–2119, https://arxiv.org/abs/1611.09187 [*DOI :* 10.1007/S10664-017-9571-8], https://hal.archives-ouvertes.fr/hal-01378523

[16] B. DANGLOT, O. L. VERA-PÉREZ, B. BAUDRY, M. MONPERRUS. *Automatic Test Improvement with DSpot: a Study with Ten Mature Open-Source Projects*, in "Empirical Software Engineering", 2018, p. 1-35, https://hal.inria.fr/hal-01923575

[17] H. GUESMI, B. BEN HEDIA, S. BLIUDZE, S. BENSALEM, B. LENABEC. *TT-BIP: using correct-by-design BIP approach for modelling real-time system with time-triggered paradigm*, in "Innovations in Systems and Software Engineering", June 2018, vol. 14, n^o 2, p. 117 - 142 [*DOI :* 10.1007/S11334-018-0312-Y], https://hal.inria.fr/hal-01823749

[18] M. MOHIUDDIN, W. SAAB, S. BLIUDZE, J.-Y. LE BOUDEC. *Axo: Detection and Recovery for Delay and Crash Faults in Real-Time Control Systems*, in "IEEE Transactions on Industrial Informatics", July 2018, vol. 14, n^o 7, p. 3065 - 3075 [*DOI :* 10.1109/TII.2017.2772219], https://hal.archives-ouvertes.fr/hal-01846124

[19] E. STACHTIARI, A. MAVRIDOU, P. KATSAROS, S. BLIUDZE, J. SIFAKIS. *Early validation of system requirements and design through correctness-by-construction*, in "Journal of Systems and Software", 2018, vol. 145, p. 52-78 [*DOI :* 10.1016/J.JSS.2018.07.053], https://hal.archives-ouvertes.fr/hal-01873999

[20] O. L. VERA-PÉREZ, B. DANGLOT, M. MONPERRUS, B. BAUDRY. *A Comprehensive Study of Pseudo-tested Methods*, in "Empirical Software Engineering", 2018, p. 1-33 [*DOI :* 10.1007/S10664-018-9653-2], https://hal.inria.fr/hal-01867423

[21] Z. YU, M. MARTINEZ, B. DANGLOT, T. DURIEUX, M. MONPERRUS. *Alleviating Patch Overfitting with Automatic Test Generation: A Study of Feasibility and Effectiveness for the Nopol Repair System*, in "Empirical Software Engineering", 2018 [*DOI :* 10.1007/S10664-018-9619-4], https://hal.inria.fr/hal-01774223

International Conferences with Proceedings

[22] S. ABITEBOUL, P. BOURHIS, V. VIANU. *Explanations and Transparency in Collaborative Workflows*, in "PODS 2018 - 37th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles Of Database Systems", Houston, Texas, United States, June 2018, https://hal.inria.fr/hal-01744978

[23] Y. AL-DHURAIBI, F. ZALILA, N. B. DJARALLAH, P. MERLE. *Coordinating Vertical Elasticity of both Containers and Virtual Machines*, in "8th International Conference on Cloud Computing and Services Science - CLOSER 2018", Funchal, Madeira, Portugal, March 2018, https://hal.archives-ouvertes.fr/hal-01683041

[24] S. CHALLITA, F. ZALILA, C. GOURDIN, P. MERLE.*A Precise Model for Google Cloud Platform*, in "6th IEEE International Conference on Cloud Engineering (IC2E)", Orlando, Florida, United States, http://conferences.computer.org/IC2E/2018/, April 2018, p. 177-183, https://hal.inria.fr/hal-01689659

[25] S. CHALLITA, F. ZALILA, P. MERLE.*Specifying Semantic Interoperability between Heterogeneous Cloud Resources with the FCLOUDS Formal Language*, in "CLOUD 2018 - 11th IEEE International Conference on Cloud Computing", San Francisco, United States, July 2018, https://hal.inria.fr/hal-01790629

[26] *Best Paper*
T. DURIEUX, Y. HAMADI, M. MONPERRUS.*Fully Automated HTML and Javascript Rewriting for Constructing a Self-healing Web Proxy*, in "Proceedings of the 29th IEEE International Symposium on Software Reliability Engineering (ISSRE 2018)", Memphis, United States, October 2018, https://arxiv.org/abs/1803.08725 [*DOI :* 10.1109/ISSRE.2018.00012], https://hal.inria.fr/hal-01746141.

[27] T. DURIEUX, Y. HAMADI, Z. YU, B. BAUDRY, M. MONPERRUS.*Exhaustive Exploration of the Failure-oblivious Computing Search Space*, in "ICST 2018 - 11th IEEE Conference on Software Testing, Validation and Verification", Västerås, Sweden, April 2018, https://arxiv.org/abs/1710.09722 - arXiv admin note: substantial text overlap with arXiv:1603.07631 [*DOI :* 10.1109/ICST.2018.00023], https://hal.inria.fr/hal-01624988

[28] G. FIENI, R. ROUVOY, L. SEINTURIER.*Towards Sustainable Software Infrastructures for Data-Intensive Systems*, in "EuroDW 2018 - 12th EuroSys Doctoral Workshop", Porto, Portugal, April 2018, https://hal.inria.fr/hal-01784225

[29] S. HABCHI, X. BLANC, R. ROUVOY.*On Adopting Linters to Deal with Performance Concerns in Android Apps*, in "ASE18 - Proceedings of the 33rd IEEE/ACM International Conference on Automated Software Engineering", Montpellier, France, ACM Press, September 2018, vol. 11 [*DOI :* 10.1145/3238147.3238197], https://hal.inria.fr/hal-01829135

[30] F. KORTE, S. CHALLITA, Z. FAIEZ, P. MERLE, J. GRABOWSKI.*Model-Driven Configuration Management of Cloud Applications with OCCI*, in "8th International Conference on Cloud Computing and Services Science (CLOSER)", Funchal, Madeira, Portugal, March 2018, p. 100-111, https://hal.inria.fr/hal-01678945

[31] M. NASERI, N. P. BORGES JR., A. ZELLER, R. ROUVOY.*AccessiLeaks: Investigating Privacy Leaks Exposed by the Android Accessibility Service*, in "PETS 2019 - The 19th Privacy Enhancing Technologies Symposium", Stockholm, Sweden, July 2019, https://hal.inria.fr/hal-01929049

[32] W. SAAB, M. MOHIUDDIN, S. BLIUDZE, J.-Y. LE BOUDEC.*Ordering Events Based on Intentionality in Cyber-Physical Systems*, in "ICCPS 2018 - 9th ACM/IEEE International Conference on Cyber-Physical Systems", Porto, Portugal, IEEE, April 2018 [*DOI :* 10.1109/ICCPS.2018.00019], https://hal.archives-ouvertes.fr/hal-01889067

[33] V. SOBREIRA, T. DURIEUX, F. MADEIRAL, M. MONPERRUS, M. A. MAIA.*Dissection of a Bug Dataset: Anatomy of 395 Patches from Defects4J*, in "SANER 2018 - 25th edition of the IEEE International Conference on Software Analysis, Evolution and Reengineering", Campobasso, Italy, March 2018, https://arxiv.org/abs/1801.06393 [*DOI :* 10.1109/SANER.2018.8330203], https://hal.inria.fr/hal-01689684

[34] S. URLI, Z. YU, L. SEINTURIER, M. MONPERRUS.*How to Design a Program Repair Bot? Insights from the Repairnator Project*, in "40th International Conference on Software Engineering, Track Software Engineering in Practice", Gothenburg, Sweden, May 2018, p. 95-104, update for oadoi on Nov 02 2018 [*DOI :* 10.1145/3183519.3183540], https://hal.inria.fr/hal-01691496

[35] A. VASTEL, P. LAPERDRIX, W. RUDAMETKIN, R. ROUVOY.*FP-Scanner: The Privacy Implications of Browser Fingerprint Inconsistencies*, in "Proceedings of the 27th USENIX Security Symposium", Baltimore, United States, August 2018, https://hal.inria.fr/hal-01820197

[36] A. VASTEL, P. LAPERDRIX, W. RUDAMETKIN, R. ROUVOY.*FP-STALKER: Tracking Browser Fingerprint Evolutions*, in "IEEE S&P 2018 - 39th IEEE Symposium on Security and Privacy", San Francisco, United States, B. PARNO, C. KRUEGEL (editors), Proceedings of the 39th IEEE Symposium on Security and Privacy (S&P), IEEE, May 2018, p. 728-741 [*DOI :* 10.1109/SP.2018.00008], https://hal.inria.fr/hal-01652021

[37] A. VASTEL, W. RUDAMETKIN, R. ROUVOY.*FP-TESTER: Automated Testing of Browser Fingerprint Resilience*, in "IWPE 2018 - 4th International Workshop on Privacy Engineering", London, United Kingdom, Proceedings of the 4th International Workshop on Privacy Engineering (IWPE'18), April 2018, p. 1-5, https://hal.inria.fr/hal-01717158

National Conferences with Proceeding

[38] *Best Paper*
F. MADEIRAL, T. DURIEUX, V. SOBREIRA, M. MAIA.*Towards an automated approach for bug fix pattern detection*, in "VEM '18 - Proceedings of the VI Workshop on Software Visualization, Evolution and Maintenance", São Carlos, Brazil, September 2018, https://arxiv.org/abs/1807.11286 , https://hal.archives-ouvertes.fr/hal-01851813.

Conferences without Proceedings

[39] M. WHITE, M. TUFANO, M. MARTINEZ, M. MONPERRUS, D. POSHYVANYK.*Sorting and Transforming Program Repair Ingredients via Deep Learning Code Similarities*, in "IEEE International Conference on Software Analysis, Evolution and Reengineering", Hangzhou, China, 2019, https://arxiv.org/abs/1707.04742 , https://hal.archives-ouvertes.fr/hal-01581170

Research Reports

[40] C. BARILLOT, I. FIJALKOW, I. QUEINNEC, F. THEOLEYRE, H. TOUZET, M. BEAUDOUIN-LAFON, M.-P. CANI, F. CHAUMETTE, G. CONREUR, V. CORTIER, C. GAVOILLE, G. GOGNIAT, J. GOSSA, A. HERZIG, J. KRIVINE, P. LAMARRE, F. LAMNABHI-LAGARRIGUE, A.-C. LETOURNEL, L. SEINTURIER, I. TELLIER, S. TORRES, C. TRUCHET, A. TSOUKIAS, M. VERLEYSEN.*Prospective Report of the scientific council of the Information Sciences and their interactions Institute of CNRS*, CNRS, December 2018, p. 1-87, https://hal.archives-ouvertes.fr/hal-01956087

[41] X. QIN, S. BLIUDZE, E. MADELAINE, M. ZHANG.*Using SMT engine to generate Symbolic Automata - Extended version*, Inria & Université Cote d'Azur, CNRS, I3S, Sophia Antipolis, France ; inria, June 2018, n^o RR-9177, https://hal.inria.fr/hal-01823507

[42] L. SEINTURIER.*Adlet: A Java-based Architecture Description Language*, Inria Lille - Nord Europe, December 2018, n^o RR-9242, https://hal.inria.fr/hal-01964792

Other Publications

[43] A. AMARILLI, P. BOURHIS, M. MONET, P. SENELLART.*Evaluating Datalog via Tree Automata and Cycluits*, October 2018, https://arxiv.org/abs/1808.04663 - 53 pages, 61 references. Journal version of "Combined Tractability of Query Evaluation via Tree Automata and Cycluits (Extended Version)" at arXiv:1612.04203. To appear in Theory of Computing Systems, https://hal.inria.fr/hal-01891814

References in notes

[44] L. CAPRA, W. EMMERICH, C. MASCOLO.*A micro-economic approach to conflict resolution in mobile computing*, in "SIGSOFT FSE", 2002, p. 31-40

[45] 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

[46] B. CORNU.*Automatic Analysis and Repair of Exception Bugs for Java Programs*, Université de Lille, November 2015, https://tel.archives-ouvertes.fr/tel-01250092

[47] 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

[48] P. ECKERSLEY.*How Unique is Your Web Browser?*, in "Proceedings of the 10th International Conference on Privacy Enhancing Technologies", Berlin, Heidelberg, PETS'10, Springer-Verlag, 2010, p. 1–18, http://dl.acm.org/citation.cfm?id=1881151.1881152

[49] 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

[50] 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

[51] 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

[52] 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

[53] 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

[54] D. JACKSON.*Alloy: A Lightweight Object Modelling Notation*, in "ACM Transactions on Software Engineering and Methodology", April 2002, vol. 11, n^o 2, p. 256-290

[55] W. LI, S. HU, J. LI, H.-A. JACOBSEN.*Community Clustering for Distributed Publish/Subscribe Systems*, in "CLUSTER", IEEE, 2012, p. 81-89

[56] 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

[57] M. MARTINEZ.*Extraction and Analysis of Knowledge for Automatic Software Repair*, Université Lille 1, October 2014, https://hal.archives-ouvertes.fr/tel-01078911

[58] P. MERLE, J.-B. STEFANI.*A formal specification of the Fractal component model in Alloy*, Inria, November 2008, n^{o} 6721, 44, http://hal.inria.fr/inria-00338987

[59] A. NOUREDDINE.*Towards a Better Understanding of the Energy Consumption of Software Systems*, Université des Sciences et Technologie de Lille - Lille I, March 2014, https://tel.archives-ouvertes.fr/tel-00961346

[60] 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

[61] 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

[62] 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

[63] 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

[64] 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

[65] 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

[66] 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

[67] 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

[68] 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

[69] 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

[70] 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

[71] 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

[72] 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