
# Activity Report Paris 2018

Edition: 2019-03-11

# List of Inria's Research Teams

# Team ALMAnaCH

## Automatic Language Modelling and ANAlysis & Computational Humanities

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
**Paris**

THEME
**Language, Speech and Audio**

# Table of contents

# Team ALMAnaCH

*Creation of the Team: 2017 January 01*

**Keywords:**

**Computer Science and Digital Science:**

A3.2.2. - Knowledge extraction, cleaning
A3.3.2. - Data mining
A3.3.3. - Big data analysis
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A9.1. - Knowledge
A9.2. - Machine learning
A9.4. - Natural language processing
A9.7. - AI algorithmics

**Other Research Topics and Application Domains:**

B1.2.2. - Cognitive science
B1.2.3. - Computational neurosciences
B9.1.1. - E-learning, MOOC
B9.5.6. - Data science
B9.6.5. - Sociology
B9.6.6. - Archeology, History
B9.6.8. - Linguistics
B9.6.10. - Digital humanities
B9.7. - Knowledge dissemination
B9.7.1. - Open access
B9.7.2. - Open data
B9.8. - Reproducibility

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Pierre Boullier [Inria, Emeritus]
Laurent Romary [Inria, Senior Researcher, HDR]
Benoît Sagot [Team leader, Inria, Researcher, HDR]
Djamé Seddah [Inria (détachement), Researcher, from Feb 2018]
Tommaso Venturini [Inria, Advanced Research Position, from Sep 2018 until Oct 2018, HDR]
Éric Villemonte de La Clergerie [Inria, Researcher]

**Faculty Member**

Djamé Seddah [Univ Paris-Sorbonne, Associate Professor, until Jan 2018]

**Technical Staff**

Wigdan Abbas Mekki Medeni [Inria, until Apr 2018]

Achraf Azhar [Inria]
Alix Chagué [CNRS (LARHRA), from Oct 2018]
Elias Benaïssa [Inria, until Apr 2018]
Farah Essaidi [Inria, Nov 2018]
Luca Foppiano [Inria]
Ganesh Jawahar [Inria, from Mar 2018]
Tanti Kristanti [Inria]
Alba Marina Malaga Sabogal [Inria]
Benjamin Muller [Inria, from Apr 2018 until Sep 2018]
Marie Puren [Inria, until Aug 2018]
Charles Riondet [Inria]
Dorian Seillier [Inria]
Lionel Tadonfouet [Inria]
Emilia Verzeni [Inria, until Apr 2018]

**PhD Students**

Jack Bowers [Vienna Academy of Sciences]
Loïc Grobol [Ecole Normale Supérieure Paris]
Mohamed Khemakhem [Inria]
Louis Martin [Facebook, from Jun 2018]
Benjamin Muller [Inria, from Oct 2018]
Pedro Ortiz Suárez [Inria, from Oct 2018]
Mathilde Regnault [Ecole Normale Supérieure Paris]
Jose Rosales Nuñez [CNRS (LIMSI), from Jun 2018]

**Post-Doctoral Fellows**

Yoann Dupont [Univ d'Orléans, from Jun 2018]
Murielle Fabre [Inria, from Nov 2018]
Ilia Markov [Inria, from Jun 2018]

**Administrative Assistants**

Christelle Guiziou [Inria, until Nov 2018]
Meriem Henni [Inria, from Nov 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

The ALMAnaCH team [0] brings together specialists of a pluri-disciplinary research domain at the interface between computer science, linguistics, philology, and statistics, namely that of **natural language processing**, **computational linguistics** and **digital and computational humanities**.

**Computational linguistics** is an interdisciplinary field dealing with the computational modelling of natural language. Research in this field is driven both by the theoretical goal of understanding human language and by practical applications in **Natural Language Processing** (hereafter NLP) such as linguistic analysis (syntactic and semantic parsing, for instance), machine translation, information extraction and retrieval, human-computer dialogue. Computational linguistics and NLP, which date back at least to the early 1950s, are among the key sub-fields of **Artificial Intelligence**.

**Digital Humanities** (hereafter DH) is an interdisciplinary field that uses computer science as a source of techniques and technologies, in particular NLP, for exploring research questions in social sciences and humanities. **Computational Humanities** aims at improving the state of the art in both computer sciences (e.g. NLP) and social sciences and humanities, by involving computer science as a research field.

[0] ALMAnaCH was created as an Inria team ("équipe") on 1st January, 2017.

ALMAnaCH is a follow-up to the ALPAGE project-team, which came to an end in December 2016. ALPAGE was created in 2007 in collaboration with Paris-Diderot University and had the status of an UMR-I since 2009. This joint team involving computational linguists from Inria as well as computational linguists from Paris-Diderot University with a strong background in linguistics proved successful. However, the context is changing, with the recent emergence of digital humanities and, more importantly, of computational humanities. This presents both an opportunity and a challenge for Inria computational linguists. It provides them with new types of data on which their tools, resources and algorithms can be used and lead to new results in human sciences. Computational humanities also provide computational linguists with new and challenging research problems, which, if solved, provide new ways of addressing research questions in the humanities.

The scientific positioning of ALMAnaCH therefore extends that of ALPAGE. We remain committed to developing state-of-the-art NLP software and resources that can be used by academics and in the industry, including recent approaches based on deep learning. At the same time we continue our work on language modelling in order to provide a better understanding of languages, an objective that is reinforced and addressed in the broader context of computational humanities, with an emphasis on language evolution and, as a result, on ancient languages.

This new scientific orientation has motivated the creation of a new project-team at the crossroads between different scientific networks, and in particular:

- The École Pratique des Hautes Études, with which collaboration has already started on a number of topics related to Digital and Computational Humanities; [0]
- The Berlin Brandenburg Academy of Sciences in Berlin which hosts the national lexicographic project in Germany, funded by the German Ministry of Education and Research (BMBF)
- CNRS's Institut des Sciences de la Communication (Institute for Communication Sciences), on topics pertaining to Digital Social Sciences; [0]
- If confirmed, the PRAIRIE Institute (PaRis Artificial Intelligence Research Institute), whose goal will be to act as a catalyst for research in Artificial Intelligence and for exchanges and between academia, industry and higher education in this domain, in which NLP plays a key role.

# 3. Research Program

## 3.1. Overview and research strands

One of the main challenges in computational linguistics is **to model and to cope with language variation**. Language varies with respect to domain and genre (news wires, scientific literature, poetry, oral transcripts...),

[0]When the ALMAnaCH team was created in January 2017, two EPHE permanent members were involved: Marc Bui, Directeur d'Études Cumulant, a specialist of computational humanities and of the computational modelling of the concept of proximity, and Daniel Stökl Ben Ezra, Directeur d'Études, a specialist of digital and computational humanities, Hebrew and Aramaic language, literature, palaeography and epigraphy. Since then, discussions and joint research endeavours have been initiated, showing the great potential of such a collaboration. Joint project proposals were submitted, one of which successfully, and we plan to work on future proposals in coming months and years. Yet after extensive discussions within all members involved in the team as well as with Éric Fleury, the head of Inria Paris, and François Jouen, Dean of the Natural Sciences department at EPHE, we came together to the conclusion that ALMAnaCH was not the optimal level for setting up a large-scale collaborative environment between both institutions, as the potential for collaboration between Inria Paris and EPHE goes well beyond NLP and text-based digital humanities. Discussions on a future Framework Agreement between EPHE and Inria Paris have started, in which ALMAnaCH will play a key role. In this context, several EPHE non-permanent members are still hosted at Inria Paris, within ALMAnaCH offices, in order to ease joint collaborations.

[0]ALMAnaCH hosted Tommaso Venturini, then on a fixed-term Senior Researcher Position, in September and October 2018, in the context of his involvment in one of ALMAnaCH's projects, the SoSweet project on Twitter-based sociolinguistics. He was granted a permanent position as CNRS Chargé de Recherches at the Institut des Sciences de la Communication starting in November 2018, and we intend to further collaborate in the future.

sociolinguistic factors (age, background, education; variation attested for instance on social media), geographical factors (dialects) and other dimensions (disabilities, for instance). But language also constantly evolves over all possible time scales. [0] Addressing this variability is still an open issue for NLP. Commonly used approaches, which often rely on supervised and semi-supervised machine learning methods, require huge amounts of annotated data. They are still struggling with the high level of variability found for instance in **user-generated content** or in **non-contemporary texts**.

ALMAnaCH tackles the challenge of language variation in two complementary directions, to which we position a specific activity related to language resources:

### *3.1.1. Research strand 1*

We focus on linguistic representations that are less affected by language variation. It obviously requires us to **stay at a state-of-the-art level in key NLP tasks** such as part-of-speech tagging and (syntactic) parsing, which are core expertise domains of ALMAnaCH members. It also requires improving the **generation of semantic representations (semantic parsing)**. This also involves the **integration of both linguistic and non-linguistic contextual information** to improve automatic linguistic analysis. This is an emerging and promising line of research in NLP. We have to identify, model and take advantage of each available type of contextual information. Addressing these issues enables us to develop new lines of research related to conversational content. Applications include chatbot-based systems and improved information and knowledge extraction algorithms. We especially focus on challenging such specific data sets as domain-specific texts or historical documents, in the larger context of the development of digital humanities.

### *3.1.2. Research strand 2*

Language variation must be better understood and modelled in all its possible realisations. In this regard, we put a strong emphasis on **three types** of language variation and their mutual interaction: **sociolinguistic variation** in synchrony (including non-canonical spelling and syntax in user-generated content), **complexity-based variation** in relation with language-related disabilities, and **diachronic variation** (computational exploration of language change and language history, with a focus ranging from Old to all forms of Modern French, as well as Indo-European languages in general). In addition, the noise introduced processes such as Optical Character Recognition (OCR) and Handwritten Text Recognition (HTR) systems, especially in the context of historical documents, bears similarities with that brought by non-canonical input in user-generated content. This noise constitutes a more transverse kind of variation stemming from the way language is graphically encoded, which we call **language-encoding variation**. [0]

### *3.1.3. Research strand 3*

Language resource development is not only a technical challenge and a necessary preliminary step to create evaluation data sets for NLP systems as well as training and for machine learning models. It is also a research field in itself, which concerns, among other challenges, (i) the development of semi-automatic and automatic algorithms to speed up the work (e.g. automatic extraction of lexical information, low-resource learning for developing pre-annotation algorithms, transfer methods to leverage tools and/or resources existing for other languages, etc.) and (ii) the development of formal models to represent linguistic information is the best possible way, thus requiring expertise at least both in NLP and in typological and formal linguistics. Language resource development involves the creation of **raw corpora from original sources** as well as the (manual, semi-automatic or automatic) development of **lexical resources** and **annotated corpora**. Such endeavours are domains of expertise of the ALMAnaCH team. This research strand 3 benefits to the whole team and beyond, and both feeds and benefits from the work of the other research strands.

[0]We do not view multilinguality as a case of language variation. Yet multilinguality, a consequence of language diversity, obviously underlies many aspect of ALMAnaCH's research activities.

[0]Other types of language variation could become research topics for ALMAnaCH in the future. This could include dialectal variation (e.g. work on Arabic) as well as the study and exploitation of paraphrases in a broader context than the above-mentioned complexity-based variation.

## 3.2. Automatic Context-augmented Linguistic Analysis

This first research strand is centred around NLP technologies and some of their applications in Artificial Intelligence (AI). Core NLP tasks such as part-of-speech tagging, syntactic and semantic parsing is improved by integrating new approaches, such as (deep) neural networks, whenever relevant, while preserving and taking advantage of our expertise on symbolic and statistical system: hybridisation not only couples symbolic and statistical approaches, but neural approaches as well. AI applications are twofold, notwithstanding the impact of language variation (see the next strand): (i) information and knowledge extraction, whatever the type of input text (from financial documents to ancient, historical texts and from Twitter data to Wikipedia) and (ii) chatbots and natural language generation. In many cases, our work on these AI applications is carried out in collaboration with industrial partners (for which cf. Section 7.1). The specificities and issues caused by language variation (a text in Old French, a contemporary financial document and tweets with a non-canonical spelling cannot be processed in the same way) are addressed in the next research strand.

### *3.2.1. Context-augmented processing of natural language at all levels: morphology, syntax, semantics*

Our expertise in NLP is the outcome of more than 10 years in developing new models of analysis and accurate techniques for the full processing of any kind of language input since the early days of the Atoll project-team and the rise of linguistically informed data-driven models as put forward within the Alpage project-team.

Traditionally, a full natural language process (NLP) chain is organised as a pipeline where each stage of analysis represents a traditional linguistic field (in a *structuralism* view) from morphological analysis to purely semantic representations. The problem is that this architecture is vulnerable to error propagation and very domain sensitive: each of these stage must be compatible at the lexical and structure levels they provide. We arguably built the best performing NLP chain for French [63], [97] and one of the best for robust multilingual parsing as shown by our results in various shared tasks over the years [93], [90], [96], [21]. So we pursue our efforts on each of our components we developed: tokenisers (e.g. SxPipe), part-of-speech taggers (e.g. MElt), constituency parsers and dependency parsers (e.g. FRMG, DyALog-SR) as well as our recent neural semantic graph parsers [90].

In particular, we continue to explore the hybridisation of symbolic and statistical approaches, and extend it to neural approaches, as initiated in the context of our participation to the CoNLL 2017 multilingual parsing shared task [0] and to Extrinsic Parsing Evaluation Shared Task [0].

Fundamentally, we want to build tools that are less sensitive to variation, more easily configurable, and self-adapting. Our short-term goals is to explore techniques such as multi-task learning (cf. already [95]) to propose a joint model of tokenisation, normalisation, morphological analysis and syntactic analysis. We also explore adversarial learning, considering the drastic variation we face in parsing user-generated content and processing historical texts, both seen as noisy input that needs to be handled at training and decoding time.

While those points are fundamental, therefore necessary, if we want to build the next generation of NLP tools, we need to *push the envelop* even further by tackling the biggest current challenge in NLP: handling the context within which a speech act is taking place.

There is indeed a strong tendency in NLP to assume that each sentence is independent from its siblings sentences as well as its context of enunciation, with the obvious objective to simplify models and reduce the complexity of predictions. While this practice is already questionable when processing full-length edited documents, it becomes clearly problematic when dealing with short sentences that are noisy, full of ellipses and external references, as commonly found in User-Generated Content (UGC).

A more expressive and context-aware structural representation of a linguistic production is required to accurately model UGC. Let us consider for instance the case for Syntax-based Machine Translation of social media content, as is carried out by the ALMAnaCH-led ANR project Parsiti (PI: DS). A Facebook post may be part of a discussion thread, which may include links to external content. Such information is required

[0] We ranked 3 for UPOS tagging and 6 for dependency parsing out of 33 participants.
[0] Semantic graph parsing, evaluated on biomedical data, speech and opinion. We ranked 1 in a joint effort with the Stanford NLP team

for a complete representation of the post's context, and in turn its accurate machine translation. Even for the presumably simpler task of POS tagging of dialogue sequences, the addition of context-based features (namely information about the speaker and dialogue moves) was beneficial [72]. In the case of UGC, working across sentence boundaries was explored for instance, with limited success, by [62] for document-wise parsing and by [82] for POS tagging.

Taking the context into account requires new inference methods able to share information between sentences as well as new learning methods capable of finding out which information is to be made available, and where. Integrating contextual information at all steps of an NLP pipeline is among the main research questions addressed in this research strand. In the short term, we focus on morphological and syntactic disambiguation within close-world scenarios, as found in video games and domain-specific UGC. In the long term, we investigate the integration of linguistically motivated semantic information into joint learning models.

From a more general perspective, contexts may take many forms and require imagination to discern them, get useful data sets, and find ways to exploit them. A context may be a question associated with an answer, a rating associated with a comment (as provided by many web services), a thread of discussions (e-mails, social media, digital assistants, chatbots—on which see below–), but also meta data about some situation (such as discussions between gamers in relation with the state of the game) or multiple points of views (pictures and captions, movies and subtitles). Even if the relationship between a language production and its context is imprecise and indirect, it is still a valuable source of information, notwithstanding the need for less supervised machine learning techniques (cf. the use of LSTM neural networks by Google to automatically suggest replies to emails).

### *3.2.2. Information and knowledge extraction*

The use of local contexts as discussed above is a new and promising approach. However, a more traditional notion of global context or world knowledge remains an open question and still raises difficult issues. Indeed, many aspects of language such as ambiguities and ellipsis can only be handled using world knowledge. Linked Open Data (LODs) such as DBpedia, WordNet, BabelNet, or Framebase provide such knowledge and we plan to exploit them.

However, each specialised domain (economy, law, medicine…) exhibits its own set of concepts with associated terms. This is also true of communities (e.g. on social media), and it is even possible to find communities discussing the same topics (e.g. immigration) with very distinct vocabularies. Global LODs weakly related to language may be too general and not sufficient for a specific language variant. Following and extending previous work in ALPAGE, we put an emphasis on information acquisition from corpora, including error mining techniques in parsed corpora (to detect specific usages of a word that are missing in existing resources), terminology extraction, and word clustering.

Word clustering is of specific importance. It relies on the distributional hypothesis initially formulated by Harris, which states that words occurring in similar contexts tend to be semantically close. The latest developments of these ideas (with word2vec or GloVe) have led to the embedding of words (through vectors) in low-dimensional semantic spaces. In particular, words that are typical of several communities (see above) can be embedded in a same semantic space in order to establish mappings between them. It is also possible in such spaces to study static configurations and vector shifts with respect to variables such as time, using topological theories (such as pretopology), for instance to explore shifts in meaning over time (cf. the ANR project Profiterole concerning ancient French texts) or between communities (cf. the ANR project SoSweet). It is also worth mentioning on-going work (in computational semantics) whose goal is to combine word embeddings to embed expressions, sentences, paragraphs or even documents into semantic spaces, e.g. to explore the similarity of documents at various time periods.

Besides general knowledge about a domain, it is important to detect and keep trace of more specific pieces of information when processing a document and maintaining a context, especially about (recurring) Named Entities (persons, organisations, locations...) —something that is the focus of future work in collaboration with Patrice Lopez on named entity detection in scientific texts. Through the co-supervision of a PhD funded by the LabEx EFL (see below), we are also involved in pronominal coreference resolution (finding the referent

of pronouns). Finally, we plan to continue working on deeper syntactic representations (as initiated with the Deep Sequoia Treebank), thus paving the way towards deeper semantic representations. Such information is instrumental when looking for more precise and complete information about who does what, to whom, when and where in a document. These lines of research are motivated by the need to extract useful contextual information, but it is also worth noting their strong potential in industrial applications.

### *3.2.3. Chatbots and text generation*

Chabots have existed for years (Eliza, Loebner prize). However, they are now becoming the focus of many concrete industrial developments, with the emergence of operational conversational agents and digital assistants (such as Siri). The current approaches mostly rely on the design of scenarios associated with very partial analysis of the requests to fill expected slots and to generate canned answers.

The next generations of such systems will rely on a deeper understanding of the requests, being able to adapt to the specificities of the users, and providing less formatted answers. We believe that chatbots are an interesting and challenging playground to deploy our expertise on knowledge acquisition (to identify concepts and formulations), information extraction based on deeper syntactic representations, context-sensitive analysis (using the thread of exchanges and profile information but also external data sources), and robustness (depending on the possible users' styles).

However, this domain of application also requires working on text generation, starting with simple canned answers and progressively moving to more sophisticated and diverse ones. This work is directly related to another line of research regarding computer-aided text simplification, for which see section 3.3.4.

## 3.3. Computational Modelling of Linguistic Variation

NLP and DH tools and resources are very often developed for contemporary, edited, non-specialised texts, often based on journalistic corpora. However, such corpora are not representative of the variety of existing textual data. As a result, the performance of most NLP systems decreases, sometimes dramatically, when faced with non-contemporary, non-edited or specialised texts. Despite the existence of domain-adaptation techniques and of robust tools, for instance for social media text processing, dealing with linguistic variation is still a crucial challenge for NLP and DH.

Linguistic variation is not a monolithic phenomenon. Firstly, it can result from different types of processes, such as variation over time (diachronic variation) and variation correlated with sociological variables (sociolinguistic variation, especially on social networks). Secondly, it can affect all components of language, from spelling (languages without a normative spelling, spelling errors of all kinds and origins) to morphology/syntax (especially in diachrony, in texts from specialised domains, in social media texts) and semantics/pragmatics (again in diachrony, for instance). Finally, it can constitute a property of the data to be analysed or a feature of the data to be generated (for instance when trying to simplify texts for increasing their accessibility for disabled and/or non-native readers).

Nevertheless, despite this variability in variation, the underlying mechanisms are partly comparable. This motivates our general vision that many generic techniques could be developed and adapted to handle different types of variation. In this regard, three aspects must be kept in mind: spelling variation (human errors, OCR/HTR errors, lack of spelling conventions for some languages...), lack or scarcity of parallel data aligning "variation-affected" texts and their "standard/edited" counterpart, and the sequential nature of the problem at hand. We will therefore explore, for instance, how unsupervised or weakly-supervised techniques could be developed and feed dedicated sequence-to-sequence models. Such architectures could help develop "normalisation" tools adapted, for example, to social media texts, texts written in ancient/dialectal varieties of well-resourced languages (e.g. Old French texts), and OCR/HTR system outputs.

Nevertheless, the different types of language variation will require specific models, resources and tools. All these directions of research constitute the core of our second research strand described in this section.

### *3.3.1. Theoretical and empirical synchronic linguistics*

Permanent members involved: all

We aim to explore computational models to deal with language variation. It is important to get more insights about language in general and about the way humans apprehend it. We will do so in at least two directions, associating computational linguistics with formal and descriptive linguistics on the one hand (especially at the morphological level) and with cognitive linguistics on the other hand (especially at the syntactic level).

Recent advances in morphology rely on quantitative and computational approaches and, sometimes, on collaboration with descriptive linguists—see for instance the special issue of the *Morphology* journal on "computational methods for descriptive and theoretical morphology", edited and introduced by [60]. In this regard, ALMAnaCH members have taken part in the design of quantitative approaches to defining and measuring morphological complexity and to assess the internal structure of morphological systems (inflection classes, predictability of inflected forms...). Such studies provide valuable insights on these prominent questions in theoretical morphology. They also improve the linguistic relevance and the development speed of NLP-oriented lexicons, as also demonstrated by ALMAnaCH members. We shall therefore pursue these investigations, and orientate them towards their use in diachronic models (see section 3.3.3).

Regarding cognitive linguistics, we have the perfect opportunity with the starting ANR-NSF project "Neuro-Computational Models of Natural Language" (NCM-NL) to go in this direction, by examining potential correlations between medical imagery applied on patients listening to a reading of "Le Petit Prince" and computation models applied on the novel. A secondary prospective benefit from the project will be information about processing evolution (by the patients) along the novel, possibly due to the use of contextual information by humans.

### 3.3.2. Sociolinguistic variation

Because language is central in our social interactions, it is legitimate to ask how the rise of digital content and its tight integration in our daily life has become a factor acting on language. This is even more actual as the recent rise of novel digital services opens new areas of expression, which support new linguistic behaviours. In particular, social media such as Twitter provide channels of communication through which speakers/writers use their language in ways that differ from standard written and oral forms. The result is the emergence of new language varieties.

A very similar situation exists with regard to historical texts, especially documentary texts or graffiti but even literary texts, that do not follow standardised orthography, morphology or syntax.

However, NLP tools are designed for standard forms of language and exhibit a drastic loss of accuracy when applied to social media varieties or non-standardised historical sources. To define appropriate tools, descriptions of these varieties are needed. However, to validate such descriptions, tools are also needed. We address this chicken-and-egg problem in an interdisciplinary fashion, by working both on linguistic descriptions and on the development of NLP tools. Recently, socio-demographic variables have been shown to bear a strong impact on NLP processing tools (see for instance [68] and references therein). This is why, in a first step, jointly with researchers involved in the ANR project SoSweet (ENS Lyon and Inria project-team Dante), we will study how these variables can be factored out by our models and, in a second step, how they can be accurately predicted from sources lacking these kinds of featured descriptions.

### 3.3.3. Diachronic variation

Language change is a type of variation pertaining to the diachronic axis. Yet any language change, whatever its nature (phonetic, syntactic...), results from a particular case of synchronic variation (competing phonetic realisations, competing syntactic constructions...). The articulation of diachronic and synchronic variation is influenced to a large extent by both language-internal factors (i.e. generalisation of context-specific facts) and/or external factors (determined by social class, register, domain, and other types of variation).

Very few computational models of language change have been developed. Simple deterministic finite-state-based phonetic evolution models have been used in different contexts. The PIElexicon project [78] uses such models to automatically generate forms attested in (classical) Indo-European languages but is based on an idiosyncratic and unacceptable reconstruction of the Proto-Indo-European language. Probabilistic finite-state

models have also been used for automatic cognate detection and proto-form reconstruction, for example by [61] and [69]. Such models rely on a good understanding of the phonetic evolution of the languages at hand.

In ALMAnaCH, our goal is to work on modelling phonetic, morphological and lexical diachronic evolution, with an emphasis on computational etymological research and on the computational modelling of the evolution of morphological systems (morphological grammar and morphological lexicon). These efforts will be in direct interaction with sub-strand 3b (development of lexical resources). We want to go beyond the above-mentioned purely phonetic models of language and lexicon evolution, as they fail to take into account a number of crucial dimensions, among which: (1) spelling, spelling variation and the relationship between spelling and phonetics; (2) synchronic variation (geographical, genre-related, etc.); (3) morphology, especially through intra-paradigmatic and inter-paradigmatic analogical levelling phenomena, (4) lexical creation, including via affixal derivation, back-formation processes and borrowings.

We apply our models to two main tasks. The first task, as developed for example in the context of the ANR project Profiterole, consists in predicting non-attested or non-documented words at a certain date based on attestations of older or newer stages of the same word (e.g., predicting a non-documented Middle French word based on its Vulgar Latin and Old French predecessors and its Modern French successor). Morphological models and lexical diachronic evolution models will provide independent ways to perform the same predictions, thus reinforcing our hypotheses or pointing to new challenges.

The second application task is computational etymology and proto-language reconstruction. Our lexical diachronic evolution models will be paired with semantic resources (wordnets, word embeddings, and other corpus-based statistical information). This will allow us to formally validate or suggest etymological or cognate relations between lexical entries from different languages of a same language family, provided they are all inherited. Such an approach could also be adapted to include the automatic detection of borrowings from one language to another (e.g. for studying the non-inherited layers in the Ancient Greek lexicon). In the longer term, we will investigate the feasibility of the automatic (unsupervised) acquisition of phonetic change models, especially when provided with lexical data for numerous languages from the same language family.

These lines of research will rely on etymological data sets and standards for representing etymological information (see Section 3.4.2).

Diachronic evolution also applies to syntax, and in the context of the ANR project Profiterole, we are beginning to explore more or less automatic ways of detecting these evolutions and suggest modifications, relying on fine-grained syntactic descriptions (as provided by meta-grammars), unsupervised sentence clustering (generalising previous works on error mining, cf. [6]), and constraint relaxation (in meta-grammar classes). The underlying idea is that a new syntactic construction evolves from a more ancient one by small, iterative modifications, for instance by changing word order, adding or deleting functional words, etc.

### *3.3.4. Accessibility-related variation*

Language variation does not always pertain to the textual input of NLP tools. It can also be characterised by their intended output. This is the perspective from which we investigate the issue of text simplification (for a recent survey, see for instance [94]). Text simplification is an important task for improving the accessibility to information, for instance for people suffering from disabilities and for non-native speakers learning a given language [79]. To this end, guidelines have been developed to help writing documents that are easier to read and understand, such as the FALC ("Facile À Lire et à Comprendre") guidelines for French. [0]

Fully automated text simplification is not suitable for producing high-quality simplified texts. Besides, the involvement of disabled people in the production of simplified texts plays an important social role. Therefore, following previous works [67], [88], our goal will be to develop tools for the computer-aided simplification of textual documents, especially administrative documents. Many of the FALC guidelines can only be linguistically expressed using complex, syntactic constraints, and the amount of available "parallel" data (aligned raw and simplified documents) is limited. We will therefore investigate hybrid techniques involving

[0]Please click here for an archived version of these guidelines (at the time this footnote is begin written, the original link does not seem to work any more).

rule-based, statistical and neural approaches based on parsing results (for an example of previous parsing-based work, see [58]). Lexical simplification, another aspect of text simplification [73], [80], will also be pursued. In this regard, we have already started a collaboration with Facebook's AI Research in Paris, the UNAPEI (the largest French federation of associations defending and supporting people with intellectual disabilities and their families), and the French Secretariat of State in charge of Disabled Persons.

Accessibility can also be related to the various presentation forms of a document. This is the context in which we have initiated the OPALINE project, funded by the *Programme d'Investissement d'Avenir - Fonds pour la Société Numérique*. The objective is for us to further develop the GROBID text-extraction suite [0] in order to be able to re-publish existing books or dictionaries, available in PDF, in a format that is accessible by visually impaired persons.

## 3.4. Modelling and Development of Language Resources

Language resources (raw and annotated corpora, lexical resources, etc.) are required in order to apply any machine learning technique (statistical, neural, hybrid) to an NLP problem, as well as to evaluate the output of an NLP system.

In data-driven, machine-learning-based approaches, language resources are the place where linguistic information is stored, be it implicitly (as in raw corpora) or explicitly (as in annotated corpora and in most lexical resources). Whenever linguistic information is provided explicitly, it complies to guidelines that formally define which linguistic information should be encoded, and how. Designing linguistically meaningful and computationally exploitable ways to encode linguistic information within language resources constitutes the first main scientific challenge in language resource development. It requires a strong expertise on both the linguistic issues underlying the type of resource under development (e.g. on syntax when developing a treebank) and the NLP algorithms that will make use of such information.

The other main challenge regarding language resource development is a consequence of the fact that it is a costly, often tedious task. ALMAnaCH members have a long track record of language resource development, including by hiring, training and supervising dedicated annotators. But a manual annotation can be speeded up by automatic techniques. ALMAnaCH members have also work on such techniques, and published work on approaches such as automatic lexical information extraction, annotation transfer from a language to closely related languages, and more generally on the use of pre-annotation tools for treebank development and on the impact of such tools on annotation speed and quality. These techniques are often also relevant for Research strand 1. For example, adapting parsers from one language to the other or developing parsers that work on more than one language (e.g. a non-lexicalised parser trained on the concatenation of treebanks from different languages in the same language family) can both improve parsing results on low-resource languages and speed up treebank development for such languages.

### *3.4.1. Construction, management and automatic annotation of Text Corpora*

Corpus creation and management (including automatic annotation) is often a time-consuming and technically challenging task. In many cases, it also raises scientific issues related for instance with linguistic questions (what is the elementary unit in a text?) as well as computer-science challenges (for instance when OCR or HTR are involved). It is therefore necessary to design a work-flow that makes it possible to deal with data collections, even if they are initially available as photos, scans, wikipedia dumps, etc.

These challenges are particularly relevant when dealing with ancient languages or scripts where fonts, OCR techniques, language models may be not extant or of inferior quality, as a result, among others, of the variety of writing systems and the lack of textual data. We will therefore work on improving print OCR for some of these languages, especially by moving towards joint OCR and language models. Of course, contemporary texts can be often gathered in very large volumes, as we already do within the ANR project SoSweet, resulting in different, specific issues.

[0] https://github.com/kermitt2/grobid

ALMAnaCH pays a specific attention to the re-usability [0] of all resources produced and maintained within its various projects and research activities. To this end, we will ensure maximum compatibility with available international standards for representing textual sources and their annotations. More precisely we will take the TEI (*Text Encoding Initiative*) guidelines as well the standards produced by ISO committee TC 37/SC 4 as essential points of reference.

From our ongoing projects in the field of Digital Humanities and emerging initiatives in this field, we observe a real need for complete but easy work-flows for exploiting corpora, starting from a set of raw documents and reaching the level where one can browse the main concepts and entities, explore their relationship, extract specific pieces of information, always with the ability to return to (fragments of) the original documents. The pieces of information extracted from the corpora also need to be represented as knowledge databases (for instance as RDF "linked data"), published and linked with other existing databases (for instance for people and locations).

The process may be seen as progressively enriching the documents with new layers of annotations produced by various NLP modules and possibly validated by users, preferably in a collaborative way. It relies on the use of clearly identified representation formats for the annotations, as advocated within ISO TC 37/SC 4 standards and the TEI guidelines, but also on the existence of well-designed collaborative interfaces for browsing, querying, visualisation, and validation. ALMAnaCH has been or is working on several of the NLP bricks needed for setting such a work-flow, and has a solid expertise in the issues related to standardisation (of documents and annotations). However, putting all these elements in a unified work-flow that is simple to deploy and configure remains to be done. In particular, work-flow and interface should maybe not be dissociated, in the sense that the work-flow should be easily piloted and configured from the interface. An option will be to identify pertinent emerging platforms in DH (such as Transkribus) and to propose collaborations to ensure that NLP modules can be easily integrated.

It should be noted that such work-flows have actually a large potential besides DH, for instance for exploiting internal documentation (for a company) or exploring existing relationships between entities.

### *3.4.2. Development of Lexical Resources*

ALPAGE, the Inriapredecessor of ALMAnaCH, has put a strong emphasis in the development of morphological, syntactic and wordnet-like semantic lexical resources for French as well as other languages (see for instance [5], [1]). Such resources play a crucial role in all NLP tools, as has been proven among other tasks for POS tagging [86], [83], [96] and parsing, and some of the lexical resource development will be targeted towards the improvement of NLP tools. They will also play a central role for studying diachrony in the lexicon, for example for Ancient to Contemporary French in the context of the Profiterole project. They will also be one of the primary sources of linguistic information for augmenting language models used in OCR systems for ancient scripts, and will allow us to develop automatic annotation tools (e.g. POS taggers) for low-resourced languages (see already [98]), especially ancient languages. Finally, semantic lexicons such as wordnets will play a crucial role in assessing lexical similarity and automating etymological research.

Therefore, an important effort towards the development of new morphological lexicons will be initiated, with a focus on ancient languages of interest. Following previous work by ALMAnaCH members, we will try and leverage all existing resources whenever possible such as electronic dictionaries, OCRised dictionaries, both modern and ancient [85], [70], [87], while using and developing (semi)automatic lexical information extraction techniques based on existing corpora [84], [89]. A new line of research will be to integrate the diachronic axis by linking lexicons that are in diachronic relation with one another thanks to phonetic and morphological change laws (e.g. XIIth century French with XVth century French and contemporary French). Another novelty will be the integration of etymological information in these lexical resources, which requires the formalisation, the standardisation, and the extraction of etymological information from OCRised dictionaries or other electronic resources, as well as the automatic generation of candidate etymologies. These directions of research are already investigated in ALMAnaCH [70], [87].

[0] From a larger point of view we intend to comply with the so-called FAIR principles (http://force11.org/group/fairgroup/fairprinciples).

An underlying effort for this research will be to further the development of the GROBID-dictionaries software, which provides cascading CRF (Conditional Random Fields) models for the segmentation and analysis of existing print dictionaries. The first results we have obtained have allowed us to set up specific collaborations to improve our performances in the domains of a) recent general purpose dictionaries such as the Petit Larousse (Nénufar project, funded by the DGLFLF in collaboration with the University of Montpellier), b) etymological dictionaries (in collaboration with the Berlin Brandenburg Academy of sciences) and c) patrimonial dictionaries such as the Dictionnaire Universel de Basnage (an ANR project, including a PhD thesis at ALMAnaCH, has recently started on this topic in collaboration with the University of Grenoble-Alpes and the University Sorbonne Nouvelle in Paris).

In the same way as we signalled the importance of standards for the representation of interoperable corpora and their annotations, we will keep making the best use of the existing standardisation background for the representation of our various lexical resources. There again, the TEI guidelines play a central role, and we have recently participated in the "TEI Lex 0" initiative to provide a reference subset for the "Dictionary" chapter of the guidelines. We are also responsible, as project leader, of the edition of the new part 4 of the ISO standard 24613 (LMF, Lexical Markup Framework) dedicated to the definition of the TEI serialisation of the LMF model (defined in ISO 24613 part 1 'Core model', 2 'Machine Readable Dictionaries' and 3 'Etymology'). We consider that contributing to standards allows us to stabilise our knowledge and transfer our competence.

### *3.4.3. Development of Annotated Corpora*

Along with the creation of lexical resources, ALMAnaCH is also involved in the creation of corpora either fully manually annotated (gold standard) or automatically annotated with state-of-the-art pipeline processing chains (silver standard). Annotations will either be only morphosyntactic or will cover more complex linguistic levels (constituency and/or dependency syntax, deep syntax, maybe semantics). Former members of the ALPAGE project have a renowned experience in those aspects (see for instance [92], [81], [91], [76]) and will participate to the creation of valuable resources originating from the historical domain genre.

Under the auspices of the ANR Parsiti project, led by ALMAnaCH (PI: DS), we aim to explore the interaction of extra-linguistic context and speech acts. Exploiting extra-linguistics context highlights the benefits of expanding the scope of current NLP tools beyond unit boundaries. Such information can be of spatial and temporal nature, for instance. They have been shown to improve Entity Linking over social media streams [65]. In our case, we decided to focus on a closed world scenario in order to study context and speech acts interaction. To do so, we are developing a multimodal data set made of live sessions of a first person shooter video game (Alien vs. Predator) where we transcribed all human players interactions and face expressions streamlined with a log of all in-game events linked to the video recording of the game session, as well as the recording of the human players themselves. The in-games events are ontologically organised and enable the modelling of the extra-linguistics context with different levels of granularity. Recorded over many games sessions, we already transcribed over 2 hours of speech that will serve as a basis for exploratory work, needed for the prototyping of our context-enhanced NLP tools. In the next step of this line of work, we will focus on enriching this data set with linguistic annotations, with an emphasis on co-references resolutions and predicate argument structures. The midterm goal is to use that data set to validate a various range of approaches when facing multimodal data in a close-world environment.

# 4. Application Domains

## 4.1. Application domains for ALMAnaCH

ALMAnaCH's research areas cover Natural Language Processing (nowadays identified as a sub-domain of Artificial Intelligence) and Digital Humanities. Application domains are therefore numerous, as witnessed by ALMAnaCH's multiple academic and industrial collaborations, for which see the relevant sections. Examples of application domains for NLP include:

- Information extraction, information retrieval, text mining (ex.: opinion surveys)
- Text generation, text simplification, automatic summarisation
- Spelling correction (writing aid, post-OCR, normalisation of noisy/non-canonical texts)
- Machine translation, computer-aided translation
- Chatbots, conversational agents, question answering systems
- Medical applications (early diagnosis, language-based medical monitoring...)
- Applications in linguistics (modelling languages and their evolution, sociolinguistic studies...)
- Digital humanities (exploitation of text documents, for instance in historical research)

# 5. New Software and Platforms

## 5.1. Enqi

- Author: Benoît Sagot
- Contact: Benoît Sagot

## 5.2. SYNTAX

KEYWORD: Parsing
FUNCTIONAL DESCRIPTION: Syntax system includes various deterministic and non-deterministic CFG parser generators. It includes in particular an efficient implementation of the Earley algorithm, with many original optimizations, that is used in several of Alpage's NLP tools, including the pre-processing chain Sx Pipe and the LFG deep parser SxLfg . This implementation of the Earley algorithm has been recently extended to handle probabilistic CFG (PCFG), by taking into account probabilities both during parsing (beam) and after parsing (n-best computation).

- Participants: Benoît Sagot and Pierre Boullier
- Contact: Pierre Boullier
- URL: http://syntax.gforge.inria.fr/

## 5.3. FRMG

KEYWORDS: Parsing - French
FUNCTIONAL DESCRIPTION: FRMG is a large-coverage linguistic meta-grammar of French. It can be compiled (using MGCOMP) into a Tree Adjoining Grammar, which, in turn, can be compiled (using DyALog) into a parser for French.

- Participant: Éric Villemonte De La Clergerie
- Contact: Éric De La Clergerie
- URL: http://mgkit.gforge.inria.fr/

## 5.4. MElt

*Maximum-Entropy lexicon-aware tagger*
KEYWORD: Part-of-speech tagger
FUNCTIONAL DESCRIPTION: MElt is a freely available (LGPL) state-of-the-art sequence labeller that is meant to be trained on both an annotated corpus and an external lexicon. It was developed by Pascal Denis and Benoît Sagot within the Alpage team, a joint Inria and Université Paris-Diderot team in Paris, France. MElt allows for using multiclass Maximum-Entropy Markov models (MEMMs) or multiclass perceptrons (multitrons) as underlying statistical devices. Its output is in the Brown format (one sentence per line, each sentence being a space-separated sequence of annotated words in the word/tag format).

MElt has been trained on various annotated corpora, using Alexina lexicons as source of lexical information. As a result, models for French, English, Spanish and Italian are included in the MElt package.

MElt also includes a normalization wrapper aimed at helping processing noisy text, such as user-generated data retrieved on the web. This wrapper is only available for French and English. It was used for parsing web data for both English and French, respectively during the SANCL shared task (Google Web Bank) and for developing the French Social Media Bank (Facebook, twitter and blog data).

- Contact: Benoît Sagot
- URL: https://team.inria.fr/almanach/melt/

## 5.5. dyalog-sr

KEYWORDS: Parsing - Deep learning - Natural language processing
FUNCTIONAL DESCRIPTION: DyALog-SR is a transition-based dependency parser, built on top of DyALog system. Parsing relies on dynamic programming techniques to handle beams. Supervised learning exploit a perceptron and aggressive early updates. DyALog-SR can handle word lattice and produce dependency graphs (instead of basic trees). It was tested during several shared tasks (SPMRL'2013 and SEMEVAL'2014). It achieves very good accuracy on French TreeBank, alone or by coupling with FRMG parser. In 2017, DyALog-SR has been extended into DyALog-SRNN by adding deep neuronal layers implemented with the Dynet library. The new version has participated to the evaluation campaigns CONLL UD 2017 (on more than 50 languages) and EPE 2017.

- Contact: Éric De La Clergerie

## 5.6. Crapbank

*French Social Media Bank*
KEYWORDS: Treebank - User-generated content
FUNCTIONAL DESCRIPTION: The French Social Media Bank is a treebank of French sentences coming from various social media sources (Twitter(c), Facebook(c)) and web forums (JeuxVidéos.com(c), Doctissimo.fr(c)). It contains different kind of linguistic annotations: - part-of-speech tags - surface syntactic representations (phrase-based representations) as well as normalized form whenever necessary.

- Contact: Djamé Seddah

## 5.7. DyALog

KEYWORD: Logic programming
FUNCTIONAL DESCRIPTION: DyALog provides an environment to compile and execute grammars and logic programs. It is essentially based on the notion of tabulation, i.e. of sharing computations by tabulating traces of them. DyALog is mainly used to build parsers for Natural Language Processing (NLP). It may nevertheless be used as a replacement for traditional PROLOG systems in the context of highly ambiguous applications where sub-computations can be shared.

- Participant: Éric Villemonte De La Clergerie
- Contact: Éric Villemonte De La Clergerie
- URL: http://dyalog.gforge.inria.fr/

## 5.8. SxPipe

KEYWORD: Surface text processing

SCIENTIFIC DESCRIPTION: Developed for French and for other languages, Sx Pipe includes, among others, various named entities recognition modules in raw text, a sentence segmenter and tokenizer, a spelling corrector and compound words recognizer, and an original context-free patterns recognizer, used by several specialized grammars (numbers, impersonal constructions, quotations...). It can now be augmented with modules developed during the former ANR EDyLex project for analysing unknown words, this involves in particular (i) new tools for the automatic pre-classification of unknown words (acronyms, loan words...) (ii) new morphological analysis tools, most notably automatic tools for constructional morphology (both derivational and compositional), following the results of dedicated corpus-based studies. New local grammars for detecting new types of entities and improvement of existing ones, developed in the context of the PACTE project, will soon be integrated within the standard configuration.
FUNCTIONAL DESCRIPTION: SxPipe is a modular and customizable processing chain dedicated to applying to raw corpora a cascade of surface processing steps (tokenisation, wordform detection, non-deterministic spelling correction. . . ). It is used as a preliminary step before ALMAnaCH's parsers (e.g., FRMG) and for surface processing (named entities recognition, text normalization, unknown word extraction and processing...).

- Participants: Benoît Sagot, Djamé Seddah and Éric Villemonte De La Clergerie
- Contact: Benoît Sagot
- URL: http://lingwb.gforge.inria.fr/

## 5.9. Mgwiki

KEYWORDS: Parsing - French
FUNCTIONAL DESCRIPTION: Mgwiki is a linguistic wiki that may used to discuss linguistic phenomena with the possibility to add annotated illustrative sentences. The work is essentially devoted to the construction of an instance for documenting and discussing FRMG, with the annotations of the sentences automatically provided by parsing them with FRMG. This instance also offers the possibility to parse small corpora with FRMG and an interface of visualization of the results. Large parsed corpora (like French Wikipedia or Wikisource) are also available. The parsed corpora can also be queried through the use of the DPath language.

- Participant: Éric Villemonte De La Clergerie
- Contact: Éric Villemonte De La Clergerie
- URL: http://alpage.inria.fr/frmgwiki/

## 5.10. WOLF

*WOrdnet Libre du Français (Free French Wordnet)*
KEYWORDS: WordNet - French - Semantic network - Lexical resource
FUNCTIONAL DESCRIPTION: The WOLF (Wordnet Libre du Français, Free French Wordnet) is a free semantic lexical resource (wordnet) for French.

The WOLF has been built from the Princeton WordNet (PWN) and various multilingual resources.

- Contact: Benoît Sagot
- URL: http://alpage.inria.fr/~sagot/wolf-en.html

## 5.11. vera

KEYWORD: Text mining
FUNCTIONAL DESCRIPTION: Automatic analysis of answers to open-ended questions based on NLP and statistical analysis and visualisation techniques (vera is currently restricted to employee surveys).

- Participants: Benoît Sagot and Dimitri Tcherniak
- Partner: Verbatim Analysis
- Contact: Benoît Sagot

## 5.12. Alexina

*Atelier pour les LEXiques INformatiques et leur Acquisition*
KEYWORD: Lexical resource
FUNCTIONAL DESCRIPTION: Alexina is ALMAnaCH's framework for the acquisition and modeling of morphological and syntactic lexical information. The first and most advanced lexical resource developed in this framework is the Lefff, a morphological and syntactic lexicon for French.

- Participant: Benoît Sagot
- Contact: Benoît Sagot
- URL: http://gforge.inria.fr/projects/alexina/

## 5.13. FQB

*French QuestionBank*
KEYWORD: Treebank
FUNCTIONAL DESCRIPTION: The French QuestionBanks is a corpus of around 2000 questions coming from various domains (TREC data set, French governmental organisation, NGOs, etc..) it contains different kind of annotations - morpho-syntactic ones (POS, lemmas) - surface syntaxe (phrase based and dependency structures) with long-distance dependency annotations.

The TREC part is aligned with the English QuestionBank (Judge et al, 2006).

- Contact: Djamé Seddah

## 5.14. Sequoia corpus

KEYWORD: Treebank
FUNCTIONAL DESCRIPTION: The Sequoia corpus contains French sentences, annotated with various linguistic information: - parts-of-speech - surface syntactic representations (both constituency trees and dependency trees) - deep syntactic representations (which are deep syntactic dependency graphs)

- Contact: Djamé Seddah

# 6. New Results

## 6.1. Syntax modelling and treebank development

**Participants:** Djamé Seddah, Benoît Sagot, Éric Villemonte de La Clergerie, Emilia Verzeni, Wigdan Abbas Mekki Medeni, Elias Benaïssa, Farah Essaidi, Amal Fethi.

- In 2018, members of ALMAnaCH have finalised a conversion of the biggest annotated data set for French, the French Treebank, to Universal Dependencies 2.3, the now *de facto* standard for syntactic annotations [27]. The same group was also deeply involved in a proposal co-written with others leaders of the field [25], aiming at representing morpho-syntactic ambiguities from user-generated content and morphologically-rich languages. This proposal was implemented via the development of language specific analysers and data-driven normalised lexica [26].
- As part of the ANR Parsiti project, the development of gold standards for North-African dialectal Arabic has seen great progresses and is coming to a pre-release date in the first semester of 2019. This work involved more than 24 man.months over the last 12 months and will culminate with a multi-layered corpus of about 2000 sentences that is made of user-generated content with a highly variable dialect that contains up to 36% of French words and mixed syntax with Arabic. In order to assess the quality of the translation produced by the Parsiti project, we also included a translation layer (North-African Arabic-French) as well as all expected morpho-syntactic and syntactic annotations, following the state-of-the-art in terms of annotations. Papers are currently being written and will target the main NLP conferences of early 2019.

- In parallel to the last item, we also translated to English half of the French Social Media Bank which was developed in our previous project [92]. A morpho-syntactic annotation layer was added. The crucial difficulty was to maintain a symmetry in term of style and level of languages between French user-generated content and its English counterpart. This data set is currently being used in the Parsiti project in order to evaluate the MT models currently being developed by the LIMSI partner.

## 6.2. Modeling of language variability via diachronic embeddings and extra-linguistic contextual features

**Participants:** Djamé Seddah, Benjamin Muller, Ganesh Jawahar, Benoît Sagot, Éric Villemonte de La Clergerie.

Following ALMAnaCH's participation in the 2017 CoNLL shared task on heavily multilingual dependency parsing in the *Universal Dependency* (hereafter UD) framework (we ranked 3rd/33 on part-of-speech tagging and 6th/33 on parsing), the team has taken part in the 2018 edition of the shared task. This year, most of the work was carried out by junior members of the team, for whom it was an interesting opportunity to gain experience on the development of NLP architectures and their deployment in the context of a shared task. It was also the opportunities to test new ideas.

We developed a neural dependency parser and a neural part-of-speech tagger, which we called 'ELMoLex' [21]. We augmented the deep Biaffine (BiAF) parser [64] with novel features to perform competitively: we utilize an in-domain version of ELMo features [77], which provide context-dependent word representations; we utilised disambiguated, embedded, morphosyntactic features extracted from our UD-compatible lexicons [26], which complements the existing feature set. In addition to incorporating character embeddings, ELMoLex leverages pre-trained word vectors, ELMo and morphosyntactic features (whenever available) to correctly handle rare or unknown words which are prevalent in languages with complex morphology. ELMoLex ranked 11th in terms of the Labeled Attachment Score metrics (70.64%) and the Morphology-aware LAS metrics (55.74%), and ranked 9th in terms of Bilexical dependency metric (60.70%). In an extrinsic evaluation setup, ELMoLex ranked 7th for Event Extraction, Negation Resolution tasks and 11th for Opinion Analysis task in terms of F1 score.

## 6.3. Modelling of language variability via diachronic embeddings and extra-linguistic contextual features

**Participants:** Djamé Seddah, Ganesh Jawahar, Éric Villemonte de La Clergerie, Benoît Sagot.

As part of the ANR SoSweet and the PHC Maimonide projects (in collaboration with Bar Ilan University for the latter), ALMAnaCH has invested a lot of efforts in 2018 into studying language variability (i.e. how the language evolve over time and how this evolution is tied to socio-demographic and dynamic network variables). Taking advantages of the SoSweet corpus (220 millions tweet) and of the Bar Ilan Hebrew Tweets (180M tweets) both collected over the last 5 years, we have been addressing the problem of studying semantic changes. We devised a novel attentional model, based on Bernouilli word embeddings, that are conditioned on contextual extra-linguistic (social) features such as network, spatial and socio-economic variables, which are associated with Twitter users, as well as topic-based features. We posit that these social features provide an inductive bias that is susceptible to helping our model to overcome the narrow time-span regime problem. Our extensive experiments reveal that, as a result of being less biased towards frequency cues, our proposed model was able to capture subtle semantic shifts and therefore benefits from the inclusion of a reduced set of contextual features. Our model thus fit the data better than current state-of-the-art dynamic word embedding models and therefore is a promising tool to study diachronic semantic changes over small time periods. A paper on this work is currently under review.

## 6.4. Standardisation of Natural Language data

**Participants:** Laurent Romary, Jack Bowers, Charles Riondet, Mohamed Khemakhem, Benoît Sagot, Loïc Grobol.

One essential aspect of working with human traces as they occur in digital humanities at large and in natural language processing in particular, is to be able to re-use any kind of primary content and further enrichments thereof. The central aspect of re-using such content is the development and applications of reference standards that reflect the best state of the art in the corresponding domains. In this respect, our team is particularly attentive to the existing standardisation background when both producing language resources or devloping NLP components. Furthermore, our specific leading roles in the domain of standardisation in both the Parthenos and EHRI EU projects as well as in related initiatives (TEI consortium, ISO committee TC 37, DARIAH lexical working group) has allowed to make progress along the following lines:

- Contributing to the revision of the ISO 24613 standard (Lexical Markup Framework) in the form of a multipart standard covering, for the time being, the core model (ISO 24613-1), machine readable dictionaries (ISO 24613-2), etymology (ISO 24613-3) and a TEI based serialisation (ISO 24613-4). Several members of the team have been particularly active as experts in the definition of the first two parts, which are now at publication and DIS stage respectively [0] and are co-editors of parts 3 and 4;
- Proposal for a reference TEI subset for integrating dictionary content: in the context of the DARIAH working group on lexical resources, a first release of the *TEI Lex 0*[0] was issued in September 2018 integrating the continuous work of the group over the the 2016-2018 period and already taken up by the infrastructure project ELEXIS [0] as its reference back-office format. This work is also the basis for the output format of Grobid-Dictionaries [71];
- Finalisation of the ISO proposal on reference annotation (ISO 24617-9): the team has been leading the work on the definition of the Reference Annotation Framework (RAF) [0] which is now at DIS ballot stage and already implemented in several concrete annotation projects[19], [43]. The standard is feature complete from a linguistic point of view (from simple co-reference to complex bridging anaphora phenomena) and compliant with the TEI stand-off annotation module [59] from the point of view of its implementation [66];
- Large-scale implementation of international standard for the documentation of the Mixtepec-Mixtec language (see section 6.11);
- Proposing a customisation architecture for the EAD international standard: EAD (Encoding Archival Description [0]) is used worldwide in cultural heritage institution to describe and exchange collection level information. In the context of the EHRI project, where we had to design a mechanism for integrating heterogeneous implementations of EAD-based data, we used the TEI ODD specification language to re-design and subset the international EAD specification to precisely provide interoperability conditions within the project[14];
- Release of the SSK (Standardisation Survival Kit), a generic environment for describing standards-based digital humanities research scenarios: the SSK is an online platform for describing research scenarios developed within the Parthenos project[40] and now deployed as a service hosted by the French national Huma-Num infrastructure [0]. The SSK has been developed as a completely open project [0], where the scenarios are themselves described as TEI-based representations[51], [35], [50].

[0] See the ISO/TC 37/SC 4 work current work program under https://www.iso.org/committee/297592/x/catalogue/p/0/u/1/w/0/d/0
[0] https://github.com/DARIAH-ERIC/lexicalresources
[0] https://elex.is
[0] https://www.iso.org/standard/69658.html
[0] https://en.wikipedia.org/wiki/Encoded_Archival_Description
[0] http://ssk.huma-num.fr
[0] https://github.com/ParthenosWP4/SSK

## 6.5. Entity-fishing: a generic named entity recognition and disambiguation for digital humanities projects

**Participants:** Marie Puren, Charles Riondet, Laurent Romary, Luca Foppiano, Tanti Kristanti.

Since several years (starting at the beginning of the EU Cendari project in 2012 [75]) we have been working on the provision of a generic named-entity recognition and disambiguation module (NERD) called *entity-fishing*[18] as a stable on-line service. The work we have achieved demonstrates the possible delivery of sustainable technical services as part of the development of research infrastructures for the humanities in Europe. In particular, our results contribute not only to DARIAH, the European digital research infrastructure for the arts and humanities, but also to OPERAS, the European research infrastructure for the development of open scholarly communication in the social sciences and humanities. Deployed as part of the French national infrastructure Huma-Num, the service provides an efficient state-of-the-art implementation coupled with standardised interfaces allowing easy deployment in a variety of potential digital humanities contexts. In 2018, we have specifically integrated *entity-fishing* within the H2020 HIRMEOS project where several open access publishers have used the service in their collections of published monographs as a means to enhance retrieval and access.

To this end, we have set up a common layer of services on top of several existing e-publishing platforms for Open Access monographs. The *entity extraction* task was deployed over a corpus of monographs provided by the HIRMEOS partners, with the following coverage:

- 4000 books in English and French from Open Edition Books
- 2000 titles in English and German from OAPEN
- 162 books in English from Ubiquity Press
- 765 books (606 in German, 159 in English) from the University of Göttingen

The introduction of *entity-fishing* has undergone different levels of integration. The majority of the participating publishers provided additional features in their user interface, using the data generated by *entity-fishing*, for example, as search facets for persons and locations to help users narrow down their searches and obtain more precise results.

*entity-fishing* has been developed in Java and it has been designed for fast processing on text and PDF, with relatively limited memory and to offer relatively close to state-of-the-art accuracy (as compared with other NERD systems). The accuracy f-score for disambiguation is currently between 76.5 and 89.1 on standard datasets (ACE2004, AIDA-CONLL-testb, AQUAINT, MSNBC) (Table 1) [74].

Table 1. Accuracy measures

| | ACE 2004 | AIDA CONLL-testb | AQUAINT | MSNBC |
|---|---|---|---|---|
| Priors | 83.1 | 66.1 | 80.3 | 71.1 |
| entity-fishing | 83.5 | 76.5 | **89.1** | 86.7 |
| Wikifier | 83.4 | 77.7 | 86.2 | 85.1 |
| DoSeR | **90.7** | 78.4 | 84.2 | 91.1 |
| AIDA | 81.5 | 77.4 | 53.2 | 78.2 |
| Spotlight | 71.3 | 59.3 | 71.3 | 51.1 |
| Babelfy | 56.1 | 59.2 | 65.2 | 60.7 |
| WAT | 80.0 | 84.3 | 76.8 | 77.7 |
| (Ganea & Hofmann, 2017) | 88.5 | **92.2** | 88.5 | **93.7** |

The objective, however, is to provide a generic service that has a steady throughput of 500-1000 words per second or one PDF page of a scientific article in 1-2 seconds on a medium range (4CPU, 3Gb Ram) Linux server.

From the point of view of the technical deployment itself, we have provided all the necessary components of a sustainable service:

- release and publish *entity-fishing* as open source software [0];
- deploy the service in the DARIAH infrastructure through HUMA-NUM [0];
- produce evaluation data and metrics for content validation.

## 6.6. From GROBID to GROBID-Dictionaries

**Participants:** Luca Foppiano, Mohamed Khemakhem, Laurent Romary, Pedro Ortiz Suárez, Alba Marina Malaga Sabogal.

GROBID is an open source software suite initiated in 2007 by Patrice Lopez with the purpose of extracting metadata automatically from scholarly papers available in PDF. Over the years, it has developed into a rich information extraction environment, and deployed in many Inria projects, but also national and international services, such as HAL (front-end meta-data extraction from uploaded scholarly publications). It is a central piece for our information extraction activities and we have been particularly active in 2018 in the following domains:

- General contributions to GROBID [0]:
  - Major refactoring and design improvements
  - fixes, tests, documentation and update of the pdf2xml fork for Windows
  - added and improved several models in collaboration with CERN (e.g. for the recognition of arXiv identifier)
  - Further tests on the specific case of bibliographic documents[32]
- Contribution to GROBID-Dictionaries [0]: the lexical GROBID extension has been implemented and tested on modern and multilingual dictionaries[23]. In the context of several collaborative activities, GROBID-Dictionaries has been applied on several documentary sources:
  - Early editions of the The Petit Larrousse Illustré in the context of the Nénufar project[45], [29]
  - Further experiments on etymological dictionaries from the Berlin Brandenburg Academy of Sciences
  - Experiments on entry-based documents such as manuscript catalogues (with University of Neuchâtel)[16] and the French address Directory Bottin from the end of the XIXth Century[22]

  These various experiments have been accompanied by an intense training and hand-on activity in the context in particular of the French research network CAHIERS (Huma-Num consortium), the Lexical Data Master Class and a series of workshop organised in South Africa under the auspices of a national linguistic documentation program. Finally, further alignments with the ongoing standardisation activities around TEI Lex0 and ISO 24613 (LMF) has been carried out to ensure a proper standards compliance of the generated output

The experience gained in the develoment and application of GROBID-Dictionaries has been the basis for the recently accepted ANR BASNUM project which aims at automatically structuring and enriching of the Dictionnaire universel (DU) by Antoine Furetière, in its 1701 edition rewritten by Basnage de Beauval and the doctoral work of Pedro Ortiz.

[0] http://github.com/kermitt2/nerd
[0] http://nerd.huma-num.fr/nerd/
[0] https://github.com/kermitt2/grobid
[0] https://github.com/MedKhem/grobid-dictionaries

## 6.7. Resources, models and tools for coreference resolution

**Participants:** Loïc Grobol, Éric Villemonte de La Clergerie.

This year we performed many experiments, some of them detailed in [28], targeting end-to-end coreference systems for spontaneous oral French. More precisely, for several mention-pair coreference detection models, we tried to assess their sensibility to various features of coreference chains and their viability for end-to-end systems, compared to the more recent antecedent scoring models.

Also, one of our objective being to assess the usefulness of syntactic features for coreference detection, we enriched the coreference annotations of the ANCOR corpus with both automatically produced dependency syntax annotations and improved speech transcription. All these annotation were wrapped in a TEI-compliant XML format as described in [20] (see also 6.4).

Finally, we have been working on neural architectures for coreference detection, building upon some recent state of the art techniques. They are based on embeddings for general text span and we try to make them more scalable through efficient uses of the local context but also more tunable to different document types and language variation. The base idea is to complete pre-training by training on related lower-level tasks such as entity-mention detection.

## 6.8. Computational history through information extraction from archive texts

**Participants:** Éric Villemonte de La Clergerie, Marie Puren, Charles Riondet, Alix Chagué, Marie-Laurence Bonhomme.

From two different DH projects emerged some interesting research questions related to the extraction of information from archival documents, in particular the management of the diversity of document types and structures and on the contrary the acquisition of detailed information from a regular visual structure.

In the context of the ANR TIME-US, whose goal is to reconstruct the "time-budgets" of textile workers in France (18th - early 20th centuries), we worked on the creation of a digitization workflow to acquire structured textual data from a wide range of printed and handwritten materials: professional court records (like *Prud'Hommes*), Police reports on strikes or early sociological studies such as the *Monographies de Le Play* This workflow has been presented at the ADHO DH conference in Mexico (see the presentation here: [34].The set up of this workflow is a prerequisite for further experiments and processing to extract information that can be exploited by historians, such as the relation between working tasks, the time spent by workers to perform them and the price they are paid for this time.

Another project was intiated in collaboration with the EPHE and the French National Archives, in the framework of the convention signed between Inria and the Ministry of Culture. This project is called LECTAUREP (for *LECTure AUtomatique de REPertoires*, and is aimed at extracting the information recorded in the registries of Parisian notaries, held by the National Archives. This project it at the intersection of NLP and Computer Vision because one of the main objectives is to extract information from the physical layout of the documents, presented as tables. Another issue is to be able to recognize with accuracy an important diversity of handwritten scripts. The final goal of LECTAUREP is to give access to researchers the information contained in these records, in particular the name of the persons involved in cases recorded by notaries, their addresses and the nature of the case (wills, powers of attorney, wedding contracts, etc.). An initial report has been produced (see [39]), and the project will continue in 2019 with the release of the extracted information (named entities, geolocation, typology, etc) into a structured database.

## 6.9. Discovering correlations between parser features and neurological observations

**Participants:** Éric Villemonte de La Clergerie, Murielle Fabre, Pauline Brunet.

In the context of the CRCNS international network, the ANR-NSF NCM-ML project (dubbed "*Petit Prince* project") aims to discover and explore correlations between features (or predictors) provided by NLP tools such as parsers, and fMRI data resulting from listening of the novel Le Petit Prince.

In 2018, Pauline Brunet, during her Master thesis, has worked on developing the infrastructure (scripts and formats) for the integration of the features, and the use of these features for computing correlations with fMRI data. A first set of features has been identified and collected from the novel and from its processing by ALMAnaCH tools (namely FRMG as an instance of a symbolic TAG-based parser and Dyalog-SR, as an instance of an hybrid feature-based neural-based dependency parser). A first dataset of fMRI scan was received to assess the infrastructure and get some preliminary results.

The work is now being continued with the arrival as a post-doc of Murielle Fabre (November 2018). With the expected arrival of the second half of the scans, she will explore more features, use her expertise to interpret the correlations, and guide the choice of new features to be tested. Since her arrival, she has in particular focused on Multi-Word Expressions (MWEs), in particular to be comparable with results published on the English side of the project. We have also identified several kinds of parsing architectures to test, in relation with various complexity parameters: (1) LSTM (two layers), (2) RNNG (with a partile filter), (3) Dyalog-SR et (4) FRMG (TAG).

In order to be in phase (and comparable) with our US partners, we have started to assemble two French corpora: - a small corpus for domain adaptation to children's books: it will permit the fine tuning of the different parsers to a great amount of dialogues an Q&A present in *Le Petit Prince*. - a large corpus of Contemporary French oral transcriptions and texts to calculate lexical association measures (AM) like PMI (Point-wise Mutual information) or Dice scores on the MWEs found in *Le Petit Prince*. This corpus of approx. 600 millions words represents a balanced counterpart to the American COCA corpus. [0]

Both Éric de La Clergerie and Murielle Fabre attended the annual meeting of the CRCNS network (Berkeley, June 2018).

## 6.10. Evaluating the quality of text simplification

**Participants:** Louis Martin, Benoît Sagot, Éric Villemonte de La Clergerie.

In 2018, our collaboration on text simplification with the Facebook Artificial Intelligence Research lab in Paris (in particular with Antoine Bordes) has started in practice. It has taken the form of a CIFRE PhD. In this context, in 2018, we dedicated important efforts to the problem of the evaluation of text simplification (TS) systems, which remains an open challenge. As the task has common points with machine translation (MT), TS is often evaluated using MT metrics such as BLEU. However, such metrics require high quality reference data, which is rarely available for TS. TS has the advantage over MT of being a monolingual task, which allows for direct comparisons to be made between the simplified text and its original version.

We compared multiple approaches to reference-less quality estimation of sentence-level TS systems, based on the dataset used for the QATS 2016 shared task. We distinguished three different dimensions: grammaticality, meaning preservation and simplicity. We have shown that $n$-gram-based MT metrics such as BLEU and METEOR correlate the most with human judgment of grammaticality and meaning preservation, whereas simplicity is best evaluated by basic length-based metrics [24].

## 6.11. Advances in descriptive, computational and historical linguistics

**Participants:** Benoît Sagot, Laurent Romary, Jack Bowers, Rebecca Blevins.

ALMAnaCH members have resumed their work in descriptive, computational and historical linguistics, an important way to ensure that NLP models and tools are robust to the diversity of world languages, as well as a way to apply NLP models and tools for contributing to research in linguistics. Three of 2018 advances in this regard are the following:

- In the context of the doctoral work of Jack Bowers, a first release of a global documentation of the Mixtepec-Mixtec language has been released which covers, multilayered annotated spoken and written resources as well as a reference lexical resource covering both basic word descriptions and elaborate semantic and etymological (word formation) content [13];

[0] https://corpus.byu.edu/coca/

- Work on language description and computational morphology for Romansh Tuatschin in collaboration with Géraldine Walther (Universität Zürich) was pursued, following the work published in 2017 [99]. A new interest in the quantitative, corpus-based study of code switching in this language has emerged in collaboration with Claudia Cathomas (Universität Zürich), leading to preliminary results to be published in 2019;
- We resumed our work in (classical) etymology in collaboration with Romain Garnier (Université de Limoges, Institut Universitaire de France), with a focus not only on (Ancient) Greek and its substrates, but also, more specifically, on Anatolian languages that could be amongst said substrates. In particular, we proposed that Lydian could be the source language for a number of Greek words lacking a good etymology in the literature [31], which motivated Rebecca Blevins's internship on the development of a lexicon of the Lydian language. We also published new etymological results at the (Proto-)Indo-European level [37].

## 6.12. Language resources and NLP tools for Medieval French

**Participants:** Éric Villemonte de La Clergerie, Mathilde Regnault, Benoît Sagot.

The main objectives of the ANR project "Profiterole" are to automatically annotate a large corpus of medieval French (9th-15th centuries) in dependency syntax and to provide a methodology for dealing with heterogeneous data like such a corpus (because of diachronic, dialectal, geographic, stylistic and genre-based variation, among other types of linguistic variation). To this end, we have continued previous experiments in morpho-syntactic tagging by trying to determine which parameters and which training sets are the best ones to use when annotating a new text. We explored two approaches for syntactic annotation (i.e. parsing). On the one hand, an ongoing thesis aims at adapting the FRMG metagrammar to medieval French, notably by changing the constraints on certain syntactic phenomena and relaxing the order of words. The development of the OFrLex lexicon has started within the Alexina framework, following the Lefff lexicon for contemporary French [5]. It already allowed for preliminary experiments. On the other hand, we conducted parsing experiments with neural models (DyALog's SRNN models). Note that members of the ALMAnaCH team participated in the CoNLL dependency parsing Shared Task 2018, which included an Old French dataset (see section 6.2).

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Industrial Collaborations

- **Verbatim Analysis**: this Inria start-up was co-created in 2009 by BS. It uses some of ALMAnaCH's free NLP software (SxPipe) as well as a data mining solution co-developed by BS, VERA, for processing employee surveys with a focus on answers to open-ended questions.
- **opensquare** A new Inria startup, opensquare, was co-created in December 2016 by BS with 2 senior specialists in HR consulting. opensquare designs, carries out and analyses employee surveys and offers HR consulting based on these results. It uses a new employee survey analysis tool, *enqi*, which is still under development.
- **Facebook**: A collaboration on text simplification ("français Facile À Lire et à Comprendre", FALC) is ongoing with Facebook's Parisian FAIR laboratory. In particular, a co-supervised (CIFRE) PhD thesis started in 2018 in collaboration with UNAPEI, the largest French federation of associations defending and supporting people with special needs and their families. This collaboration is expected to pave the way for a larger initiative involving (at least) these three partners as well as the relevant ministries.
- **Bluenove**: A contract with this company has been signed, which defines the framework of our collaboration on the integration of NLP tools (e.g. chatbot-related modules) within Bluenove's plateform Assembl, dedicated to online citizen debating forums. It involves a total of 24 months of fixed-term contracts (12 months for a post-doc, currently working on the project, and 12 months for a research engineer, which is still to be recruited).

- **Science Miner**: ALMAnaCH (previously ALPAGE) has been collaborating since 2014 years with this company founded by Patrice Lopez, a specialist in machine learning techniques and initiator of the GROBID and NERD (now entity-fishing) suites. Patrice Lopez provides scientific support on the corresponding software components in the context of the Parthenos, EHRI and Iperion projects, as well as in the context of the Inria anHALytics initiative, aiming at providing a scholarly dashboard on the scientific papers available from the HAL national publication repository.
- **Hyperlex** A collaboration was initiated in 2018 on NLP for legal documents, involving especially EVdLC.
- ALMAnaCH members led a proposal for the creation of an ANR LabCom with Fortia Financial Solutions, a company specialized in *Financial Technology*, namely the analysis of financial documents from investment funds. The proposal has been rejected. Meanwhile, this project is currently being extended toward a FUI with Systran, the market leader in specialized machine translation systems, and the BNP as industrial partner. The funding requested will cross the 3 millions euros bar.
- ALMAnaCH members have recently initiated discussions with other companies (Louis Vuitton, Suez...), so that additional collaborations might start in the near future. They have also presented their work to companies interested in knowing more about the activities of Inria Paris in AI and NLP.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

### *8.1.1. ANR*

- **ANR SoSweet** (2015-2019, PI J.-P. Magué, resp. ALMAnaCH: DS; Other partners: ICAR [ENS Lyon, CRNS], Dante [Inria]). Topic: studying sociolinguistic variability on Twitter, comparing linguistic and graph-based views on tweets
- **ANR ParSiTi** (2016-2021, PI Djamé Seddah, Other partners: LIMSI, LIPN). Topic: context-aware parsing and machine translation of user-generated content
- **ANR PARSE-ME** (2015-2020, PI. Matthieu Constant, resp. Marie Candito [ALPAGE, then LLF], ALMAnaCH members are associated with Paris-Diderot's LLF for this project). Topic: multi-word expressions in parsing
- **ANR Profiterole** (2016-2020, PI Sophie Prévost [LATTICE], resp. Benoit Crabbé [ALPAGE, then LLF], ALMAnaCH members are associated with Paris-Diderot's LLF for this project). Topic: modelling and analysis of Medieval French
- **ANR TIME-US** (2016-2019, PI Manuela Martini [LARHRA], ALMAnaCH members are associated with Paris-Diderot's CEDREF for this project). Topic: Digital study of remuneration and time budget textile trades in XVIIIth and XIXth century France
- **ANR BASNUM** (2018-2021, PI Geoffrey Williams [Université Grenoble Alpes], resp. ALMAnaCH: LR). Topic: Digitalisation and computational linguistic study of Basnage de Beauval's *Dictionnaire universel* published in 1701.

### *8.1.2. Competitivity Clusters*

- **LabEx EFL** (2010-2019, PI Christian Puech [HTL, Paris 3], Sorbonne Paris Cité). Topic: empirical foundations of linguistics, including computational linguistics and natural language processing. ALPAGE was one of the partner teams of this LabEx, which gathers a dozen of teams within and around Paris whose research interests include one aspects of linguistics or more. BS serves as deputy head (and former head) of one of the scientific strands of the LabEx, namely strand 6 dedicated to language resources. BS and DS are in charge of a number of scientific "operations" within strands 6, 5 ("computational semantic analysis") and 2 ("experimental grammar"). BS, EVdLC and DS are now

individual members of the LabEx EFL since 1st January 2017, and BS still serves as the deputy head of strand 6. Main collaborations are on language resource development (strands 5 and 6), syntactic and semantic parsing (strand 5, especially with LIPN [CNRS and U.Paris 13]) and computational morphology (strands 2 and 6, especially with CRLAO [CNRS and Inalco]).

### *8.1.3. Other National Initiatives*

- **LECTAUREP project** (2017-2018): An explorative study has been launched in collaboration with the National Archives in France, in the context of the framework agreement between Inria and the Ministry of Culture, to explore the possibility of extracting various components from digitized 19th Century notary registers.
- **Nénufar (DGLFLF - Délégation générale à la langue française et aux langues de France)**: The projects is intended to digitize and exploit the early editions (beginning of the 20th Century) of the Petit Larousse dictionary. ALMAnaCH is involve to contribute to the automatic extraction of the dictionary content by means of GROBID-Dictionaries and define a TEI compliant interchange format for all results.
- **PIA Opaline**: The objective of the project is to provide a better access to published French literature and reference material for visually impaired persons. Financed by the Programme d'Investissement d'Avenir, it will integrate technologies related to document analysis and re-publishing, textual content enrichment and dedicated presentational interfaces. Inria participate to deploy the GROBID tool suite for the automatic structuring of content from books available as plain PDF files.

## 8.2. European Initiatives

### *8.2.1. FP7 & H2020 Projects*

- **H2020 Parthenos** (2015-2019, PI Franco Niccolucci [University of Florence]; LR is a work package coordinator) Topic: strengthening the cohesion of research in the broad sector of Linguistic Studies, Humanities, Cultural Heritage, History, Archaeology and related fields through a thematic cluster of European Research Infrastructures, integrating initiatives, e-infrastructures and other world-class infrastructures, and building bridges between different, although tightly interrelated, fields.
- **H2020 EHRI** "European Holocaust Research Infrastructure" (2015-2019, PI Conny Kristel [NIOD-KNAW, NL]; LR is task leader) Topic: transform archival research on the Holocaust, by providing methods and tools to integrate and provide access to a wide variety of archival content.
- **H2020 Iperion CH** (2015-2019, PI Luca Pezzati [CNR, IT], LR is task leader) Topic: coordinating infrastructural activities in the cultural heritage domain.
- **H2020 HIRMEOS**: HIRMEOS objective is to improve five important publishing platforms for the open access monographs in the humanities and enhance their technical capacities and services and rendering technologies, while making their content interoperable. Inria is responsable for improving integrating the entity-fishing component deplyed as an infrastructural service for the five platforms.
- **H2020 DESIR**: The DESIR project aims at contributing to the sustainability of the DARIAH infrastructure along all its dimensions: dissemination, growth, technology, robustness, trust and education. Inria is responsable for providing of a portfolio of text analytics services based on GROBID and entity-fishing.

### *8.2.2. Collaborations in European Programs, Except FP7 & H2020*

- **ERIC DARIAH "Digital Research Infrastructure for the Arts and Humanities"** (set up as a consortium of states, 2014-2034; LR served president of the board of director until August 2018) Topic: coordinating Digital Humanities infrastructure activities in Europe (17 partners, 5 associated partners).
- **COST enCollect** (2017-2020, PI Lionel Nicolas [European Academy of Bozen/Bolzano]) Topic: combining language learning and crowdsourcing for developing language teaching materials and more generic language resources for NLP

### 8.2.3. Collaborations with Major European Organizations

Collaborations with institutions not cited above (for the SPMRL initiative, see below):

- Universität Zürich, Switzerland (Géraldine Walther) [computational morphology, lexicons]
- Berlin-Brandenburgische Akademie der Wissenschaften [Berlin-Brandenburg Academy of Sciences and Humanities], Berlin, Germany (Alexander Geyken) [lexicology]
- Österreichische Akademie der Wissenschaften [Austrian Academy of Sciences], Vienna, Austria (Karlheinz Moerth) [lexicology]
- University of Cambridge, United Kingdom (Ekaterina Kochmar) [text simplification]
- Univerza v Ljubljani [University of Ljubljana], Ljubljana, Slovenia (Darja Fišer) [wordnet development]

## 8.3. International Initiatives

### 8.3.1. Participation in International Programs

**PHC Maimonide** (2018-2019, PI Djamé Seddah, co-PI Yoav Goldberg (Bar Ilan University)). Topics: Building NLP resources for analyzing reactions to major events in Hebrew and French social media.

## 8.4. International Research Visitors

### 8.4.1. Visits of International Scientists

- Dr. Ekaterina Kochmar (University of Cambridge), 3 days in June
- Dr. Teresa Lynn (Dublin City University), 2 stays of 1 week each.

# 9. Dissemination

## 9.1. Promoting Scientific Activities

- LR was invited to present an overview of information extraction methods in the humanities in the context of the conference cycle: Ringvorlesung "Open Technology for an Open Society", Jan 2018, Berlin, Germany

### 9.1.1. Scientific Events Organisation

#### 9.1.1.1. General Chair, Scientific Chair

- LR: Co-chair of the Lexical Data Masterclass, Berlin, 3-7 December https://digilex.hypotheses.org/551
- Mohamed Khemakhem: Chair of the GROBID-Camp: Inria de Paris 27th March 2018

### 9.1.2. Scientific Events Selection

#### 9.1.2.1. Member of the Conference Program/Scientific/Reviewing Committees

- BS: Member of the Program, Scientific or Reviewing Committee of the following conferences and workshops: ACL 2018, NAACL 2018, International Morphology Meeting 2018, Int'l Colloquium on Loanwords and Substrata 2018
- LR: Member of the Program, Scientific or Reviewing Committee of the following conferences and workshops: Fourteenth Joint ACL - ISO Workshop on Interoperable Semantic Annotation, COLING 2018, TPDL 2018, ACL 2018, NAACL-HLT 2018, TOTh 2018, ELPUB 2018, DHd2018, LDL-2018, DH 2018

- DS: Member of the Program, Scientific or Reviewing Committee of the following conferences and workshops: ACL 2018, EMNLP 2018,CoNLL 2018, COLING 2018, EthicNLP 2018, LREC 2018, WNUT 2018, LAW-MWE-CxG 2018.
- EVdLC: Program Committee member and reviewer for LREC, ACL, COLING, NAACL, ToTH, EMNLP

### *9.1.3. Journal*

*9.1.3.1. Member of the Editorial Boards*

- LR: Member of the JTEI advisory board
- LR: Member of the scientific board of the Revue Humanités numériques

*9.1.3.2. Reviewer - Reviewing Activities*

- BS: Reviewer for the following journals: *Language Resource and Evaluation*, *Traitement Automatique des Langues*
- LR: Reviewer for the following journals: Language Resource and Evaluation Journal, Journal of the TEI
- DS: Reviewer for the following journals: TALLIP, LRE, NLE, Poznan Studies in Contemporary Linguistics, Computational Linguistics

### *9.1.4. Invited Talks*

- BS was invited to give a talk to Master 2 computational linguistics students and University staff at the Université Grenoble Alpes (November)
- LR was invited to give talks at Open-Access-Tage, Sep 2018, Graz; Workshop DARIAH-CH, University of Neuchâtel, November 2018; "Stay tuned to the future", an international conference on the impact of research infrastructures for social sciences and humanities – bologna, January 2018; NIMS, Tskuba, Japan, September 2018; Rétro-numérisation de documents historiques et partage dans le Web sémantique : l'exemple de la lexicographie – Atelier de formation annuel du consortium Cahier – Montpellier – 26-29 juin 2018; "Serving Learning and Scholarship", Fiesole retreat, Barcelona, April 2018
- DS was invited to give a talk at the Indiana University's department of linguistics (October), at Bar Ilan University (November) respectively on Noisy User-Generated Content Treebanking and on Tackling language variability via diachronic word embeddings.

### *9.1.5. Training*

- Mohamed Khemakhem chaired and tutored the GROBID-Dictionaries series:
  - BBAW & Praxiling joint workshop - Berlin: February 2018
  - Atelier de formation annuel du consortium Cahier – Montpellier – 26-29 June 2018
  - SADiLaR GROBID-Dictionaries Workshop (Pretoria) : October 26, 2018
  - SADiLaR GROBID-Dictionaries Workshop (Potchefstroom) : October 30, 2018 from
  - SADiLaR GROBID-Dictionaries Workshop (StellenBosch) : November 2, 2018
  - Lexical Data Masterclass 2018 - Berlin 3-7 December 2018
- Mathilde Reignault attended the ESSLLI 2018 Summer School in Language and Information as part of her doctoral studies training.

### *9.1.6. Leadership within the Scientific Community*

- LR: President of the board of directors of DARIAH (until August 2018)
- LR: Member of the board of directors or the TEI consortium
- LR: President of ISO committee TC 37 (Language and terminology)

- LR: Member of the ELEXIS Interoperability and Sustainability Committee (ISC) — ELEXIS is the European Lexicographic Infrastructure (https://elex.is)
- EVdLC: Chairman of the ACL special interest group SIGPARSE
- BS: Member, Deputy Treasurer and Member of the Board of the Société de Linguistique de Paris
- DS: Board member of the French NLP society (Atala, 2017-2020), Vice-President of the Atala and program chair of the "journée d'études".
- DS: Member of the ACL's BIG (Broad Interest Group) Diversity group.
- : Charles Riondet: Co-chair of the DARIAH Guidelines and Standards Working Group.
- : Marie Puren: Co-chair of the DARIAH Guidelines and Standards Working Group.

### *9.1.7. Scientific Expertise*

- BS: member of the recruitment committee for the new "ingénieur d'études" position in Inria Paris's communication department
- LR: has carried out various project assesment expertises for: City University Honk Kong, the go!digital programm at the Austrian Academy of Sciences, the Haifa-Technion Joint Research Submission to Milgrom Foundation, teh Swiss National Science Foundation
- DS: Project evaluation for the Flanders Research Agency.
- EVdLC: Evaluator for a European COST proprosal
- EVdLC: Evaluator for the Program Call of DGLFLF on "Langue et Numérique"

### *9.1.8. Research Administration*

- BS: Member of the Board of Inria Paris's Scientific Committee ("Comité des Projets")
- BS: Member of the International Relations Working Group of Inria's Scientific and Technological Orientation Council (COST-GTRI)
- BS: Deputy Head of the research strand on Language Resources of the LabEx EFL (Empirical Foundations of Linguistics), and is therefore a deputy member of the Governing Board of the LabEx; BS and DS are in charge of several research operations in the LabEx
- LR: President of the board of directors of DARIAH
- LR: President of the scientific committee of ABES (Agence Bibliographique de l'Enseignement Supérieur)
- LR: President of ISO committee TC 37 (Language and Terminology)
- Mohamed Khemakhem and LR: Project leaders of the ISO 24613-4 LMF "TEI Serialisation"
- LR: Convener of ISO working group TC 37/SC 4/WG 4 (lexical resources)
- LR: Member of the Text Encoding Initiative board
- LR: advisor for scientific information to the director for science at Inria

## 9.2. Teaching - Supervision - Juries

### *9.2.1. Teaching*

- Master: Benoît Sagot (with Emmanuel Dupoux), "Speech and Language Processing", 20h, M2, Master "Mathématiques, Vision, Apprentissage", ENS Paris-Saclay, France
- Licence: Djamé Seddah, "Certificat Informatique et Internet", 30h, L1-L2-L3, Université Paris Sorbonne, France
- Master: Djamé Seddah, "Modèles pour la linguistique computationnelle",36h, M1, Université Paris Sorbonne, France
- Master: Djamé Seddah, "Traduction automatique", 30h, M2, Université Paris Sorbonne, France

- Master: Loïc Grobol, “Introduction à la fouille de textes”, 39h, M1, Université Sorbonne Nouvelle, France
- Master: Yoann Dupont and Loïc Grobol, “Langages de script”, 39h, M2, INALCO, France

### *9.2.2. Supervision*

HdR: Benoît Sagot, “Informatiser le lexique — Modélisation, développement et exploitation de lexiques morphologiques, syntaxiques et sémantiques”, 28th June 2018, mentored by Laurent Romary

PhD in progress: Mohamed Khemakhem, “Structuration automatique de dictionnaires à partir de modèles lexicaux standardisés”, September 2016, Paris Diderot, supervised by Laurent Romary

PhD in progress: Loïc Grobol, ““Reconnaissance automatique de chaînes de coréférences en français par combinaison d’apprentissage automatique et de connaissances linguistiques”, “Université Sorbonne Nouvelle”, started in Oct. 2016, supervised by Frédéric Landragin (main superviser), Isabelle Tellier† (main superviser), Éric de La Clergerie and Marco Dinarelli

PhD in progress: Jack Bowers, “Technology, description and theory in language documentation: creating a comprehensive body of multi-media resources for Mixtepec-Mixtec using standards, ontology and Cognitive Linguistics”, started in Oct. 2016, EPHE, supervised by Laurent Romary

PhD in progress: Axel Herold, “Automatic identification and modeling of etymological information from retro-digitized dictionaries”, October 2016, EPHE, Laurent Romary

PhD in progress: Mathilde Regnault,“Annotation et analyse de corpus hétérogènes”, “Université Sorbonne Nouvelle”, started in Oct. 2017, supervised by Sophie Prévost (main superviser), Isabelle Tellier†, and Éric de la Clergerie

PhD in progress: Pedro Ortiz, “Automatic Enrichment of Ancient Dictionaries”, October 2018, Sorbonne Université, supervised by Laurent Romary and Benoît Sagot

PhD in progress: Benjamin Muller, “Multi-task learning for text normalisation, parsing and machine translation”, October 2018, Sorbonne Université, supervised by Benoît Sagot and Djamé Seddah

PhD in progress: José Carlos Rosales, supervised by Guillaume Wisnewski (Limsi) and Djamé Seddah

### *9.2.3. Juries*

- BS: president of the Habilitation committee for Kim Gerdes at Université Paris Nanterre on November 29th (Title: “Same Same but Different: Paradigms in Syntax”; Mentor: Sylvain Kahane)
- BS: reviewer (“rapporteur”) in the PhD committee for Sébastien Delecraz at Aix-Marseille Université on December 10th (Title: “Approches jointes texte/image pour la compréhension multimodale de documents”; Supervisor: )
- LR: member of the PhD committee for Cyrille Suire, University of La Rochelle, September 2019 (Title: "Recherche d’information et humanités numériques : une approche et des outils pour l’historien")
- BS: member of the recruiting committee for a communication officer at Inria Paris (Aug–Oct 2018)
- LR: member of the selection committee for the assistant professor position on linguistics and NLP at University of Orléans (May 2018)

## 9.3. Popularization

### *9.3.1. Interventions*

- Welcoming of schoolchildren at Inria Paris (half a day with ALMAnaCH members within an one-week-long stay; December 2018)

- ALMAnaCH members were involved in the Profiterole ANR project's presentation at the *Salon de l'Innovation* of the conference TALN 2018 (the "SiTAL" show).
- Presentation in Education Network ISN "Informatique et Science du Numérique" (March)
- Invited speaker in a citizen debate on Artificial Intelligence (association "Les coteaux en Seine", Bougival, November 21st 2018)

# 10. Bibliography

## Major publications by the team in recent years

[1] D. FIŠER, B. SAGOT.*Constructing a poor man's wordnet in a resource-rich world*, in "Language Resources and Evaluation", 2015, vol. 49, n° 3, p. 601-635 [*DOI :* 10.1007/s10579-015-9295-6], https://hal.inria.fr/hal-01174492

[2] P. LOPEZ, L. ROMARY.*HUMB: Automatic Key Term Extraction from Scientific Articles in GROBID*, in "SemEval 2010 Workshop", Uppsala, Sweden, ACL SigLex event, July 2010, 4, https://hal.inria.fr/inria-00493437

[3] C. RIBEYRE, É. VILLEMONTE DE LA CLERGERIE, D. SEDDAH.*Because Syntax does Matter: Improving Predicate-Argument Structures Parsing Using Syntactic Features*, in "Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies", Denver, USA, United States, Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, June 2015, https://hal.archives-ouvertes.fr/hal-01174533

[4] L. ROMARY.*TEI and LMF crosswalks*, in "JLCL - Journal for Language Technology and Computational Linguistics", 2015, vol. 30, n° 1, https://hal.inria.fr/hal-00762664

[5] B. SAGOT.*The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French*, in "7th international conference on Language Resources and Evaluation (LREC 2010)", Valletta, Malta, May 2010, https://hal.inria.fr/inria-00521242

[6] B. SAGOT, É. VILLEMONTE DE LA CLERGERIE.*Error Mining in Parsing Results*, in "Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics", Sydney, Australia, Association for Computational Linguistics, July 2006, p. 329–336

[7] D. SEDDAH, B. SAGOT, M. CANDITO, V. MOUILLERON, V. COMBET.*The French Social Media Bank: a Treebank of Noisy User Generated Content*, in "COLING 2012 - 24th International Conference on Computational Linguistics", Mumbai, Inde, Kay, Martin and Boitet, Christian, December 2012, http://hal.inria.fr/hal-00780895

[8] R. TSARFATY, D. SEDDAH, Y. GOLDBERG, S. KÜBLER, Y. VERSLEY, M. CANDITO, J. FOSTER, I. REHBEIN, L. TOUNSI.*Statistical Parsing of Morphologically Rich Languages (SPMRL) What, How and Whither*, in "Proceedings of the NAACL HLT 2010 First Workshop on Statistical Parsing of Morphologically-Rich Languages", États-Unis Los Angeles, Association for Computational Linguistics, 2010, p. 1–12

[9] R. TSARFATY, D. SEDDAH, S. KÜBLER, J. NIVRE.*Parsing Morphologically Rich Languages: Introduction to the Special Issue*, in "Computational Linguistics", March 2013, vol. 39, n$^{o}$ 1, 8 [*DOI :* 10.1162/COLI_A_00133], https://hal.inria.fr/hal-00780897

[10] É. VILLEMONTE DE LA CLERGERIE.*Improving a symbolic parser through partially supervised learning*, in "The 13th International Conference on Parsing Technologies (IWPT)", Naria, Japan, November 2013, https://hal.inria.fr/hal-00879358

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] B. SAGOT.*Computerising the lexicon : Modelling, development and use of morphological, syntactic and semantic lexicons*, Sorbonne Université, June 2018, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01895229

### Articles in International Peer-Reviewed Journal

[12] S. BENIAMINE, O. BONAMI, B. SAGOT.*Inferring inflection classes with description length*, in "Journal of Language Modelling", February 2018, vol. 5, n$^{o}$ 3, p. 465–525, https://hal.inria.fr/hal-01718879

[13] J. BOWERS, L. ROMARY.*Bridging the Gaps between Digital Humanities, Lexicography, and Linguistics: A TEI Dictionary for the Documentation of Mixtepec-Mixtec*, in "Dictionaries: Journal of the Dictionary Society of North America", 2018, vol. 39, n$^{o}$ 2, p. 79-106, https://hal.inria.fr/hal-01968871

[14] L. ROMARY, C. RIONDET.*EAD-ODD: A solution for project-specific EAD schemes*, in "Archival Science", April 2018, Special thanks to Annelies van Nispen (NIOD) and Hector Martinez Alonso (ALMAnaCH) for their help, and to Lou Burnard (TEI) for his wise comments [*DOI :* 10.1007/s10502-018-9290-y], https://hal.inria.fr/hal-01737568

### Invited Conferences

[15] A. BERTINO, L. FOPPIANO, L. ROMARY, P. MOUNIER.*Leveraging Concepts in Open Access Publications*, in "PUBMET 2018 - 5th Conference on Scholarly Publishing in the Context of Open Science", Zadar, Croatia, September 2018, https://hal.inria.fr/hal-01900303

[16] M. KHEMAKHEM, L. ROMARY, S. GABAY, H. BOHBOT, F. FRONTINI, G. LUXARDO.*Automatically Encoding Encyclopedic-like Resources in TEI*, in "The annual TEI Conference and Members Meeting", Tokyo, Japan, September 2018, https://hal.inria.fr/hal-01819505

### International Conferences with Proceedings

[17] J. BOWERS, L. ROMARY.*Encoding Mixtepec-Mixtec Etymology in TEI*, in "TEI Conference and Members' Meeting", Tokyo, Japan, September 2018, https://hal.inria.fr/hal-02003975

[18] L. FOPPIANO, L. ROMARY.*entity-fishing: a DARIAH entity recognition and disambiguation service*, in "Digital Scholarship in the Humanities", Tokyo, Japan, September 2018, https://hal.inria.fr/hal-01812100

[19] L. GROBOL, F. LANDRAGIN, S. HEIDEN.*XML-TEI-URS: using a TEI format for annotated linguistic resources*, in "CLARIN Annual Conference 2018", Pisa, Italy, October 2018, https://hal.archives-ouvertes.fr/hal-01827563

[20] L. GROBOL, I. TELLIER, É. VILLEMONTE DE LA CLERGERIE, M. DINARELLI, F. LANDRAGIN.*ANCOR-AS: Enriching the ANCOR Corpus with Syntactic Annotations*, in "LREC 2018 - 11th edition of the Language Resources and Evaluation Conference", Miyazaki, Japan, May 2018, https://hal.inria.fr/hal-01744572

[21] G. JAWAHAR, B. MULLER, A. FETHI, L. MARTIN, É. VILLEMONTE DE LA CLERGERIE, B. SAGOT, D. SEDDAH.*ELMoLex: Connecting ELMo and Lexicon features for Dependency Parsing*, in "CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies", Brussels, Belgium, October 2018 [*DOI :* 10.18653/v1/K18-2023], https://hal.inria.fr/hal-01959045

[22] M. KHEMAKHEM, C. BRANDO, L. ROMARY, F. MÉLANIE-BECQUET, J.-L. PINOL.*Fueling Time Machine: Information Extraction from Retro-Digitised Address Directories*, in "JADH2018 "Leveraging Open Data"", Tokyo, Japan, September 2018, https://hal.archives-ouvertes.fr/hal-01814189

[23] M. KHEMAKHEM, A. HEROLD, L. ROMARY.*Enhancing Usability for Automatically Structuring Digitised Dictionaries*, in "GLOBALEX workshop at LREC 2018", Miyazaki, Japan, May 2018, https://hal.archives-ouvertes.fr/hal-01708137

[24] L. MARTIN, S. HUMEAU, P.-E. MAZARÉ, A. BORDES, É. VILLEMONTE DE LA CLERGERIE, B. SAGOT.*Reference-less Quality Estimation of Text Simplification Systems*, in "1st Workshop on Automatic Text Adaptation (ATA)", Tilburg, Netherlands, November 2018, https://arxiv.org/abs/1901.10746 , https://hal.inria.fr/hal-01959054

[25] A. MORE, Ö. ÇETINOĞLU, Ç. ÇÖLTEKIN, N. HABASH, B. SAGOT, D. SEDDAH, D. TAJI, R. TSARFATY.*CoNLL-UL: Universal Morphological Lattices for Universal Dependency Parsing*, in "11th Language Resources and Evaluation Conference", Miyazaki, Japan, May 2018, https://hal.inria.fr/hal-01786125

[26] B. SAGOT.*A multilingual collection of CoNLL-U-compatible morphological lexicons*, in "Eleventh International Conference on Language Resources and Evaluation (LREC 2018)", Miyazaki, Japan, May 2018, https://hal.inria.fr/hal-01798798

[27] D. SEDDAH, É. VILLEMONTE DE LA CLERGERIE, B. SAGOT, H. MARTINEZ ALONSO, M. CANDITO.*Cheating a Parser to Death: Data-driven Cross-Treebank Annotation Transfer*, in "Eleventh International Conference on Language Resources and Evaluation (LREC 2018)", Miyazaki, Japan, May 2018, https://hal.inria.fr/hal-01798801

### National Conferences with Proceeding

[28] M. BRASSIER, A. PURET, A. VOISIN-MARRAS, L. GROBOL.*Classification par paires de mention pour la résolution des coréférences en français parlé interactif*, in "Conférence jointe CORIA-TALN-RJC 2018", Rennes, France, ATALA and ARIA, May 2018, https://hal.inria.fr/hal-01821213

### Conferences without Proceedings

[29] H. BOHBOT, F. FRONTINI, G. LUXARDO, M. KHEMAKHEM, L. ROMARY.*Presenting the Nénufar Project: a Diachronic Digital Edition of the Petit Larousse Illustré*, in "GLOBALEX 2018 - Globalex workshop at LREC2018", Miyazaki, Japan, May 2018, p. 1-6, https://hal.archives-ouvertes.fr/hal-01728328

[30] M. DINARELLI, L. GROBOL.*Modeling a label global context for sequence tagging in recurrent neural networks*, in "Journée commune AFIA-ATALA sur le Traitement Automatique des Langues et l'Intelligence

Artificielle pendant la onzième édition de la plate-forme Intelligence Artificielle (PFIA 2018)", Nancy, France, July 2018, https://hal.archives-ouvertes.fr/hal-02002111

[31] R. GARNIER, B. SAGOT.*New results on a centum substratum in Greek: the Lydian connection*, in "International Colloquium on Loanwords and Substrata in Indo-European languages", Limoges, France, June 2018, https://hal.inria.fr/hal-01798979

[32] D. LINDEMANN, M. KHEMAKHEM, L. ROMARY.*Retro-digitizing and Automatically Structuring a Large Bibliography Collection*, in "European Association for Digital Humanities (EADH) Conference", Galway, Ireland, EADH, December 2018, https://hal.archives-ouvertes.fr/hal-01941534

[33] H. MARAOUI, K. HADDAR, L. ROMARY.*Segmentation tool for hadith corpus to generate TEI encoding*, in "4th International Conference on Advanced Intelligent Systems and Informatics (AISI'18)", Cairo, Egypt, September 2018, https://hal.archives-ouvertes.fr/hal-01794105

[34] M. PUREN, A. CHAGUÉ, M. MARTINI, É. VILLEMONTE DE LA CLERGERIE, C. RIONDET.*Creating gold data to understand the gender gap in the French textile trades (17th–20th century). Time-Us project*, in "Digital Humanities 2018: "Puentes/ Bridges"", Mexico, Mexico, June 2018, https://hal.archives-ouvertes.fr/hal-01850080

[35] M. PUREN, C. RIONDET, L. ROMARY, D. SEILLIER, L. TADJOU.*The Standardization Survival Kit (SSK): Bringing best practices to research communities in the Humanities*, in "Digital Humanities Benelux 2018", Amsterdam, Netherlands, June 2018, https://hal.archives-ouvertes.fr/hal-01850075

[36] M. PUREN, D. SEILLIER, C. RIONDET, L. TADJOU.*Le Standardization Survival Kit (SSK): Faciliter l'usage des standards dans les Humanités*, in "Rencontres de la TGIR Huma-Num", Ecully, France, June 2018, https://hal.archives-ouvertes.fr/hal-01850078

[37] B. SAGOT.*A new PIE root *h1er '(to be) dark red, dusk red': drawing the line between inherited and borrowed words for 'red(ish)', 'pea', 'ore', 'dusk' and 'love' in daughter languages*, in "International Colloquium on Loanwords and Substrata in Indo-European languages", Limoges, France, June 2018, https://hal.inria.fr/hal-01798976

### Scientific Books (or Scientific Book chapters)

[38] T. BLANKE, C. KRISTEL, L. ROMARY.*Crowds for Clouds: Recent Trends in Humanities Research Infrastructures*, in "Cultural Heritage Digital Tools and Infrastructures", A. BENARDOU, E. CHAMPIO, C. DALLAS, L. HUGHES (editors), Routledge, 2018, https://arxiv.org/abs/1601.00533 , https://hal.inria.fr/hal-01248562

### Research Reports

[39] M.-L. BONHOMME.*Répertoire des Notaires parisiens Segmentation automatique et reconnaissance d'écriture : Rapport exploratoire*, Inria, October 2018, p. 1-10, https://hal.inria.fr/hal-01949198

[40] C. RIONDET, D. SEILLIER, L. TADJOU, L. ROMARY.*Standardization Survival Kit (Final)*, Inria Paris, October 2018, https://hal.inria.fr/hal-01925144

### Scientific Popularization

[41] M. PUREN, C. RIONDET, D. SEILLIER, L. TADJOU.*The Standardization Survival Kit : For a wider use of standards within Arts and Humanities*, April 2018, Journée de formation : "Gérer et explorer les données textuelles", https://hal.inria.fr/hal-01763688

[42] C. RIONDET.*TEI: de l'image au texte : Décrire son corpus grâce aux métadonnées*, February 2018, p. 1-69, TEI: de l'image au texte, https://hal.inria.fr/hal-01708839

## Other Publications

[43] A. ADLI, E. ENGEL, L. ROMARY, F. SAME.*A stand-off XML-TEI representation of reference annotation*, March 2018, DGfS 2018: 40. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft, Poster, https://hal.inria.fr/hal-01876327

[44] A. BERTINO, L. FOPPIANO, L. ROMARY, P. MOUNIER.*Leveraging Concepts in Open Access Publications*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01981922

[45] H. BOHBOT, A. FAUCHERE, F. FRONTINI, A. JACKIEWICZ, G. LUXARDO, A. STEUCKARDT, M. KHEMAKHEM, L. ROMARY.*A Diachronic Digital Edition of the Petit Larousse illustré*, May 2018, Journée d'étude CORLI : Traitements et standardisation des corpus multimodaux et web 2.0., Poster, https://hal.archives-ouvertes.fr/hal-01873805

[46] J. BOWERS.*Language Documentation and Standards in Digital Humanities: TEI and the documentation of Mixtepec-Mixtec*, February 2019, working paper or preprint, https://hal.inria.fr/hal-02004005

[47] S. GABAY, M. KHEMAKHEM, L. ROMARY.*GROBID and catalogues*, November 2018, Lecture, https://hal.archives-ouvertes.fr/cel-01951107

[48] K. ILLMAYER, M. PUREN.*How to work together successfully with e-Humanities and e-Heritage Research Infrastructures (PARTHENOS Webinar)*, February 2018, Lecture, https://hal.archives-ouvertes.fr/cel-01731455

[49] F. LANDRAGIN, M. DELABORDE, Y. DUPONT, L. GROBOL.*Description et modélisation des chaînes de référence. Le projet ANR Democrat (2016-2020) et ses avancées à mi-parcours*, May 2018, Cinquième édition du Salon de l'Innovation en TAL (Traitement Automatique des Langues) et RI (Recherche d'Informations), Poster, https://hal.archives-ouvertes.fr/hal-01797982

[50] M. PUREN, C. RIONDET, L. ROMARY, D. SEILLIER, L. TADJOU.*SSK by example. Make your Arts and Humanities research go standard*, June 2018, Digital Humanities 2018 : "Bridges/Puentes", Poster, https://hal.archives-ouvertes.fr/hal-01848882

[51] M. PUREN, C. RIONDET, L. ROMARY, D. SEILLIER, L. TADJOU.*The SSK. Make your Arts and Humanities research go standard. TEI inside !*, September 2018, TEI2018 - Annual TEI Conference and Members Meeting, Poster, https://hal.inria.fr/hal-01902702

[52] D. REINEKE, L. ROMARY.*Reference SKOS and TBX vocabularies used for comparing the two standards*, September 2018, technical document associated to a journal submission, https://hal.inria.fr/hal-01883377

[53] C. RIONDET.*Stewardship of Cultural Heritage Data. In the shoes of a researcher*, April 2018, Cultural Heritage Data Re-use Charter Feedback workshop hosted by the LIBER Digital Humanities & Digital Cultural Heritage Working group, https://hal.inria.fr/hal-01762295

[54] C. RIONDET.*Traces de l'héroïsme. Le programme mémoriel de la résistance parisienne*, February 2018, À paraître en 2018 dans "La clandestinité politique, des anarchistes au djihadisme (XXe-XXIe siècle)" (sous la dir. de Cirefice, France, Le Quang, Riondet), https://hal.inria.fr/hal-01715006

[55] C. RIONDET.*À la recherche de l'archive clandestine*, February 2018, À paraître en 2018 dans "La clandestinité politique, des anarchistes au djihadisme (XXe-XXIe siècle)" (sous la dir. de Cirefice, France, Le Quang, Riondet), https://hal.inria.fr/hal-01715002

[56] L. ROMARY.*Data Mining Technologies at the service of Open Knowledge*, January 2018, p. 1-65, Ringvorlesung "Open Technology for an Open Society", https://hal.inria.fr/hal-01708771

[57] L. ROMARY.*Open Access in France: how the call of Jussieu reflects our social, technical and political landscape*, September 2018, Open-Access-Tage, https://hal.inria.fr/hal-01881469

## References in notes

[58] M. J. ARANZABE, A. D. DE ILARRAZA, I. GONZALEZ-DIOS.*Transforming complex sentences using dependency trees for automatic text simplification in Basque*, in "Procesamiento del lenguaje natural", 2013, vol. 50, p. 61–68

[59] P. BANSKI, B. GAIFFE, P. LOPEZ, S. MEONI, L. ROMARY, T. SCHMIDT, P. STADLER, A. WITT.*Wake up, standOff!*, September 2016, TEI Conference 2016, https://hal.inria.fr/hal-01374102

[60] O. BONAMI, B. SAGOT.*Computational methods for descriptive and theoretical morphology: a brief introduction*, in "Morphology", 2017, vol. 27, n° 4, p. 1-7 [*DOI :* 10.1017/CBO9781139248860], https://hal.inria.fr/hal-01628253

[61] A. BOUCHARD-CÔTÉ, D. HALL, T. GRIFFITHS, D. KLEIN.*Automated Reconstruction of Ancient Languages using Probabilistic Models of Sound Change*, in "Proceedings of the National Academy of Sciences", 2013, n° 110, p. 4224–4229

[62] J. C. K. CHEUNG, G. PENN.*Utilizing Extra-sentential Context for Parsing*, in "Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing", Cambridge, Massachusetts, EMNLP '10, 2010, p. 23–33

[63] M. CONSTANT, M. CANDITO, D. SEDDAH.*The LIGM-Alpage Architecture for the SPMRL 2013 Shared Task: Multiword Expression Analysis and Dependency Parsing*, in "Fourth Workshop on Statistical Parsing of Morphologically Rich Languages", Seattle, United States, October 2013, p. 46-52, https://hal.archives-ouvertes.fr/hal-00932372

[64] T. DOZAT, C. D. MANNING.*Deep Biaffine Attention for Neural Dependency Parsing*, in "CoRR", 2016, vol. abs/1611.01734, http://arxiv.org/abs/1611.01734

[65] Y. FANG, M. CHANG.*Entity Linking on Microblogs with Spatial and Temporal Signals*, in "TACL", 2014, vol. 2, p. 259–272, https://tacl2013.cs.columbia.edu/ojs/index.php/tacl/article/view/323

[66] L. GROBOL, F. LANDRAGIN, S. HEIDEN.*Interoperable annotation of (co)references in the Democrat project*, in "Thirteenth Joint ISO-ACL Workshop on Interoperable Semantic Annotation", Montpellier, France, H.

BUNT (editor), ACL Special Interest Group on Computational Semantics (SIGSEM) and ISO TC 37/SC 4 (Language Resources) WG 2, September 2017, https://hal.archives-ouvertes.fr/hal-01583527

[67] J. E. HOARD, R. WOJCIK, K. HOLZHAUSER.*An automated grammar and style checker for writers of Simplified English*, in "Computers and Writing: State of the Art", 1992, p. 278–296

[68] D. HOVY, T. FORNACIARI.*Increasing In-Class Similarity by Retrofitting Embeddings with Demographic Information*, in "Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing", Association for Computational Linguistics, 2018, p. 671–677, http://aclweb.org/anthology/D18-1070

[69] D. HRUSCHKA, S. BRANFORD, E. SMITH, J. WILKINS, A. MEADE, M. PAGEL, T. BHATTACHARYA.*Detecting Regular Sound Changes in Linguistics as Events of Concerted Evolution*, in "Current Biology", 2015, vol. 1, n$^{o}$ 25, p. 1–9

[70] M. KHEMAKHEM, L. FOPPIANO, L. ROMARY.*Automatic Extraction of TEI Structures in Digitized Lexical Resources using Conditional Random Fields*, in "electronic lexicography, eLex 2017", Leiden, Netherlands, September 2017, https://hal.archives-ouvertes.fr/hal-01508868

[71] M. KHEMAKHEM, L. FOPPIANO, L. ROMARY.*Automatic Extraction of TEI Structures in Digitized Lexical Resources using Conditional Random Fields*, in "electronic lexicography, eLex 2017", Leiden, Netherlands, September 2017, https://hal.archives-ouvertes.fr/hal-01508868

[72] S. KÜBLER, M. SCHEUTZ, E. BAUCOM, R. ISRAEL.*Adding Context Information to Part Of Speech Tagging for Dialogues*, in "NEALT Proceedings Series", M. DICKINSON, K. MUURISEP, M. PASSAROTTI (editors), 2010, vol. 9, p. 115-126

[73] A.-L. LIGOZAT, C. GROUIN, A. GARCIA-FERNANDEZ, D. BERNHARD.*Approches à base de fréquences pour la simplification lexicale*, in "TALN-RÉCITAL 2013", 2013, 493

[74] P. LOPEZ.*entity-fishing*, in "WikiDATA Conf", September 2017, https://grobid.s3.amazonaws.com/presentations/29-10-2017.pdf

[75] P. LOPEZ, A. MEYER, L. ROMARY.*CENDARI Virtual Research Environment & Named Entity Recognition techniques*, February 2014, Grenzen überschreiten – Digitale Geisteswissenschaft heute und morgen, Poster, https://hal.inria.fr/hal-01577975

[76] H. MARTINEZ ALONSO, D. SEDDAH, B. SAGOT.*From Noisy Questions to Minecraft Texts: Annotation Challenges in Extreme Syntax Scenarios*, in "2nd Workshop on Noisy User-generated Text (W-NUT) at CoLing 2016", Osaka, Japan, December 2016, https://hal.inria.fr/hal-01584054

[77] M. E. PETERS, M. NEUMANN, M. IYYER, M. GARDNER, C. CLARK, K. LEE, L. ZETTLEMOYER.*Deep contextualized word representations*, in "Proc. of NAACL", 2018

[78] J. PYSSALO.*System PIE: the Primary Phoneme Inventory and Sound Law System for Proto-Indo-European*, University of Helsinki, 2013

[79] L. RELLO, R. BAEZA-YATES, S. BOTT, H. SAGGION.*Simplify or help?: text simplification strategies for people with dyslexia*, in "Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility", ACM, 2013, 15

[80] L. RELLO, R. BAEZA-YATES, L. DEMPERE-MARCO, H. SAGGION.*Frequent words improve readability and short words improve understandability for people with dyslexia*, in "IFIP Conference on Human-Computer Interaction", Springer, 2013, p. 203–219

[81] C. RIBEYRE, M. CANDITO, D. SEDDAH.*Semi-Automatic Deep Syntactic Annotations of the French Treebank*, in "The 13th International Workshop on Treebanks and Linguistic Theories (TLT13)", Tübingen, Germany, Proceedings of TLT 13, Tübingen Universität, December 2014, https://hal.inria.fr/hal-01089198

[82] A. M. RUSH, R. REICHART, M. COLLINS, A. GLOBERSON.*Improved Parsing and POS Tagging Using Inter-sentence Consistency Constraints*, in "Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning", Jeju Island, Korea, EMNLP-CoNLL '12, 2012, p. 1434–1444

[83] B. SAGOT, H. MARTINEZ ALONSO.*Improving neural tagging with lexical information*, in "15th International Conference on Parsing Technologies", Pisa, Italy, September 2017, p. 25-31, https://hal.inria.fr/hal-01592055

[84] B. SAGOT, D. NOUVEL, V. MOUILLERON, M. BARANES.*Extension dynamique de lexiques morphologiques pour le français à partir d'un flux textuel*, in "TALN - Traitement Automatique du Langage Naturel", Les sables d'Olonne, France, June 2013, p. 407-420, https://hal.inria.fr/hal-00832078

[85] B. SAGOT.*DeLex, a freely-avaible, large-scale and linguistically grounded morphological lexicon for German*, in "Language Resources and Evaluation Conference", Reykjavik, Iceland, European Language Resources Association, May 2014, https://hal.inria.fr/hal-01022288

[86] B. SAGOT.*External Lexical Information for Multilingual Part-of-Speech Tagging*, Inria Paris, June 2016, n° RR-8924, https://hal.inria.fr/hal-01330301

[87] B. SAGOT.*Extracting an Etymological Database from Wiktionary*, in "Electronic Lexicography in the 21st century (eLex 2017)", Leiden, Netherlands, September 2017, p. 716-728, https://hal.inria.fr/hal-01592061

[88] C. SCARTON, M. DE OLIVEIRA, A. CANDIDO JR, C. GASPERIN, S. M. ALUÍSIO.*SIMPLIFICA: a tool for authoring simplified texts in Brazilian Portuguese guided by readability assessments*, in "Proceedings of the NAACL HLT 2010 Demonstration Session", Association for Computational Linguistics, 2010, p. 41–44

[89] Y. SCHERRER, B. SAGOT.*A language-independent and fully unsupervised approach to lexicon induction and part-of-speech tagging for closely related languages*, in "Language Resources and Evaluation Conference", Reykjavik, Iceland, European Language Resources Association, May 2014, https://hal.inria.fr/hal-01022298

[90] S. SCHUSTER, É. VILLEMONTE DE LA CLERGERIE, M. D. CANDITO, B. SAGOT, C. D. MANNING, D. SEDDAH.*Paris and Stanford at EPE 2017: Downstream Evaluation of Graph-based Dependency Representations*, in " EPE 2017 - The First Shared Task on Extrinsic Parser Evaluation", Pisa, Italy, Proceedings of the 2017 Shared Task on Extrinsic Parser Evaluation, September 2017, p. 47-59, https://hal.inria.fr/hal-01592051

[91] D. SEDDAH, M. CANDITO.*Hard Time Parsing Questions: Building a QuestionBank for French*, in "Tenth International Conference on Language Resources and Evaluation (LREC 2016)", Portorož, Slovenia, Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC 2016), May 2016, https://hal.archives-ouvertes.fr/hal-01457184

[92] D. SEDDAH, B. SAGOT, M. CANDITO, V. MOUILLERON, V. COMBET.*The French Social Media Bank: a Treebank of Noisy User Generated Content*, in "COLING 2012 - 24th International Conference on Computational Linguistics", Mumbai, India, Kay, Martin and Boitet, Christian, December 2012, https://hal.inria.fr/hal-00780895

[93] D. SEDDAH, B. SAGOT, M. CANDITO.*The Alpage Architecture at the SANCL 2012 Shared Task: Robust Pre-Processing and Lexical Bridging for User-Generated Content Parsing*, in "SANCL 2012 - First Workshop on Syntactic Analysis of Non-Canonical Language , an NAACL-HLT'12 workshop", Montréal, Canada, June 2012, https://hal.inria.fr/hal-00703124

[94] M. SHARDLOW.*A survey of automated text simplification*, in "International Journal of Advanced Computer Science and Applications", 2014, vol. 4, n$^{o}$ 1, p. 58–70

[95] A. SØGAARD, Y. GOLDBERG.*Deep multi-task learning with low level tasks supervised at lower layers*, in "Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)", Berlin, Germany, 2016, p. 231–235

[96] É. VILLEMONTE DE LA CLERGERIE, B. SAGOT, D. SEDDAH.*The ParisNLP entry at the ConLL UD Shared Task 2017: A Tale of a #ParsingTragedy*, in "Conference on Computational Natural Language Learning", Vancouver, Canada, Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, August 2017, p. 243-252 [*DOI :* 10.18653/v1/K17-3026], https://hal.inria.fr/hal-01584168

[97] É. VILLEMONTE DE LA CLERGERIE.*Jouer avec des analyseurs syntaxiques*, in "TALN 2014", Marseilles, France, ATALA, July 2014, https://hal.inria.fr/hal-01005477

[98] G. WALTHER, B. SAGOT.*Speeding up corpus development for linguistic research: language documentation and acquisition in Romansh Tuatschin*, in "Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature", Vancouver, Canada, Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, August 2017, p. 89 - 94 [*DOI :* 10.18653/v1/W17-2212], https://hal.inria.fr/hal-01570614

[99] G. WALTHER, B. SAGOT.*Speeding up corpus development for linguistic research: language documentation and acquisition in Romansh Tuatschin*, in "Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature", Vancouver, Canada, Proceedings of the Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature, August 2017, p. 89 - 94 [*DOI :* 10.18653/v1/W17-2212], https://hal.inria.fr/hal-01570614

# Project-Team ALPINES

## Algorithms and parallel tools for integrated numerical simulations

IN COLLABORATION WITH: Laboratoire Jacques-Louis Lions (LJLL)

IN PARTNERSHIP WITH:
**CNRS**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Distributed and High Performance Computing**

# Table of contents

## Project-Team ALPINES

*Creation of the Team: 2013 January 01, updated into Project-Team: 2014 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.5. - Numerical Linear Algebra
A6.2.7. - High performance computing
A6.3. - Computation-data interaction
A6.3.1. - Inverse problems
A7.1. - Algorithms

**Other Research Topics and Application Domains:**

B3.3.1. - Earth and subsoil
B9.5.2. - Mathematics
B9.5.3. - Physics

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Laura Grigori [Team leader, Inria, Senior Researcher, HDR]
Sever Hirstoaga [Inria, Researcher, from Dec 2018]
Frédéric Nataf [CNRS, Senior Researcher]

**Faculty Members**

Xavier Claeys [Univ Pierre et Marie Curie, Associate Professor]
Frédéric Hecht [Univ Pierre et Marie Curie, Professor]

**Technical Staff**

Simplice Donfack [Inria, until Oct 2018]
Axel Fourmont [Inria, until Sep 2018]
Franck Houssen [Inria, until Mar 2018]
Pierre-Henri Tournier [CNRS]

**PhD Students**

Hussam Al Daas [Inria]
Alan Ayala Obregon [Inria, until Sep 2018]
Igor Chollet [Univ Pierre et Marie Curie]
Thibault Cimic [Inria, from Oct 2018]
Zakariae Jorti [IFPEN]
Pierre Marchand [Inria]
Van Thanh Nguyen [Inria]
Olivier Tissot [Inria]

**Post-Doctoral Fellows**

Jan Papez [Inria]
Amin Rafiei [Campus France, until May 2018]

**Visiting Scientists**

Agnieszka Miedlar [University of Kansas, from Jun 2018 until Jul 2018]

Qiang Niu [Xi'an Jiaotong Liverpool University, from May 2018 until Jul 2018]

**Administrative Assistant**

Laurence Bourcier [Inria]

# 2. Overall Objectives

## 2.1. Introduction

The focus of our research is on the development of novel parallel numerical algorithms and tools appropriate for state-of-the-art mathematical models used in complex scientific applications, and in particular numerical simulations. The proposed research program is by nature multi-disciplinary, interweaving aspects of applied mathematics, computer science, as well as those of several specific applications, as porous media flows, elasticity, wave propagation in multi-scale media.

Our first objective is to develop numerical methods and tools for complex scientific and industrial applications, that will enhance their scalable execution on the emergent heterogeneous hierarchical models of massively parallel machines. Our second objective is to integrate the novel numerical algorithms into a middle-layer that will hide as much as possible the complexity of massively parallel machines from the users of these machines.

# 3. Research Program

## 3.1. Overview

The research described here is directly relevant to several steps of the numerical simulation chain. Given a numerical simulation that was expressed as a set of differential equations, our research focuses on mesh generation methods for parallel computation, novel numerical algorithms for linear algebra, as well as algorithms and tools for their efficient and scalable implementation on high performance computers. The validation and the exploitation of the results is performed with collaborators from applications and is based on the usage of existing tools. In summary, the topics studied in our group are the following:

- Numerical methods and algorithms
  - Mesh generation for parallel computation
  - Solvers for numerical linear algebra
  - Computational kernels for numerical linear algebra
- Validation on numerical simulations

## 3.2. Domain specific language - parallel FreeFem++

In the engineering, researchers, and teachers communities, there is a strong demand for simulation frameworks that are simple to install and use, efficient, sustainable, and that solve efficiently and accurately complex problems for which there are no dedicated tools or codes available. In our group we develop FreeFem++ (see http://www.freefem.org/ff++), a user dedicated language for solving PDEs. The goal of FreeFem++ is not to be a substitute for complex numerical codes, but rather to provide an efficient and relatively generic tool for:

- getting a quick answer to a specific problem,
- prototyping the resolution of a new complex problem.

The current users of FreeFem++ are mathematicians, engineers, university professors, and students. In general for these users the installation of public libraries as MPI, MUMPS, Ipopt, Blas, lapack, OpenGL, fftw, scotch, is a very difficult problem. For this reason, the authors of FreeFem++ have created a user friendly language, and over years have enriched its capabilities and provided tools for compiling FreeFem++ such that the users do not need to have special knowledge of computer science. This leads to an important work on porting the software on different emerging architectures.

Today, the main components of parallel FreeFem++ are:

1. definition of a coarse grid,
2. splitting of the coarse grid,
3. mesh generation of all subdomains of the coarse grid, and construction of parallel data structures for vectors and sparse matrices from the mesh of the subdomain,
4. call to a linear solver,
5. analysis of the result.

All these components are parallel, except for point (5) which is not in the focus of our research. However for the moment, the parallel mesh generation algorithm is very simple and not sufficient, for example it addresses only polygonal geometries. Having a better parallel mesh generation algorithm is one of the goals of our project. In addition, in the current version of FreeFem++, the parallelism is not hidden from the user, it is done through direct calls to MPI. Our goal is also to hide all the MPI calls in the specific language part of FreeFem++.

## 3.3. Solvers for numerical linear algebra

Iterative methods are widely used in industrial applications, and preconditioning is the most important research subject here. Our research considers domain decomposition methods and iterative methods and its goal is to develop solvers that are suitable for parallelism and that exploit the fact that the matrices are arising from the discretization of a system of PDEs on unstructured grids.

One of the main challenges that we address is the lack of robustness and scalability of existing methods as incomplete LU factorizations or Schwarz-based approaches, for which the number of iterations increases significantly with the problem size or with the number of processors. This is often due to the presence of several low frequency modes that hinder the convergence of the iterative method. To address this problem, we study different approaches for dealing with the low frequency modes as coarse space correction in domain decomposition or deflation techniques.

We also focus on developing boundary integral equation methods that would be adapted to the simulation of wave propagation in complex physical situations, and that would lend themselves to the use of parallel architectures. The final objective is to bring the state of the art on boundary integral equations closer to contemporary industrial needs. From this perspective, we investigate domain decomposition strategies in conjunction with boundary element method as well as acceleration techniques (H-matrices, FMM and the like) that would appear relevant in multi-material and/or multi-domain configurations. Our work on this topic also includes numerical implementation on large scale problems, which appears as a challenge due to the peculiarities of boundary integral equations.

## 3.4. Computational kernels for numerical linear algebra

The design of new numerical methods that are robust and that have well proven convergence properties is one of the challenges addressed in Alpines. Another important challenge is the design of parallel algorithms for the novel numerical methods and the underlying building blocks from numerical linear algebra. The goal is to enable their efficient execution on a diverse set of node architectures and their scaling to emerging high-performance clusters with an increasing number of nodes.

Increased communication cost is one of the main challenges in high performance computing that we address in our research by investigating algorithms that minimize communication, as communication avoiding algorithms. We propose to integrate the minimization of communication into the algorithmic design of numerical linear algebra problems. This is different from previous approaches where the communication problem was addressed as a scheduling or as a tuning problem. The communication avoiding algorithmic design is an aproach originally developed in our group since 2007 (initially in collaboration with researchers from UC Berkeley and CU Denver). While at mid term we focus on reducing communication in numerical linear algebra, at long term we aim at considering the communication problem one level higher, during the parallel mesh generation tool described earlier.

# 4. Application Domains

## 4.1. Compositional multiphase Darcy flow in heterogeneous porous media

We study the simulation of compositional multiphase flow in porous media with different types of applications, and we focus in particular on reservoir/bassin modeling, and geological CO2 underground storage. All these simulations are linearized using Newton approach, and at each time step and each Newton step, a linear system needs to be solved, which is the most expensive part of the simulation. This application leads to some of the difficult problems to be solved by iterative methods. This is because the linear systems arising in multiphase porous media flow simulations cumulate many difficulties. These systems are non-symmetric, involve several unknowns of different nature per grid cell, display strong or very strong heterogeneities and anisotropies, and change during the simulation. Many researchers focus on these simulations, and many innovative techniques for solving linear systems have been introduced while studying these simulations, as for example the nested factorization [Appleyard and Cheshire, 1983, SPE Symposium on Reservoir Simulation].

## 4.2. Inverse problems

We focus on methods related to the blend of time reversal techniques and absorbing boundary conditions (ABC) used in a non standard way. Since the seminal paper by [M. Fink et al., Imaging through inhomogeneous media using time reversal mirrors. Ultrasonic Imaging, 13(2):199, 1991.], time reversal is a subject of very active research. The principle is to back-propagate signals to the sources that emitted them. The initial experiment was to refocus, very precisely, a recorded signal after passing through a barrier consisting of randomly distributed metal rods. In [de Rosny and Fink. Overcoming the difraction limit in wave physics using a time-reversal mirror and a novel acoustic sink. Phys. Rev. Lett., 89 (12), 2002], the source that created the signal is time reversed in order to have a perfect time reversal experiment. In [41], we improve this result from a numerical point of view by showing that it can be done numerically without knowing the source. This is done at the expense of not being able to recover the signal in the vicinity of the source. In [42], time dependent wave splitting is performed using ABC and time reversal techniques. We now work on extending these methods to non uniform media.

All our numerical simulations are performed in FreeFem++ which is very flexible. As a byproduct, it enables us to have an end user point of view with respect to FreeFem++ which is very useful for improving it.

## 4.3. Numerical methods for wave propagation in multi-scale media

We are interested in the development of fast numerical methods for the simulation of electromagnetic waves in multi-scale situations where the geometry of the medium of propagation may be described through caracteristic lengths that are, in some places, much smaller than the average wavelength. In this context, we propose to develop numerical algorithms that rely on simplified models obtained by means of asymptotic analysis applied to the problem under consideration.

Here we focus on situations involving boundary layers and *localized* singular perturbation problems where wave propagation takes place in media whose geometry or material caracteristics are submitted to a small scale perturbation localized around a point, or a surface, or a line, but not distributed over a volumic sub-region of the propagation medium. Although a huge literature is already available for the study of localized singular perturbations and boundary layer pheneomena, very few works have proposed efficient numerical methods that rely on asymptotic modeling. This is due to their functional framework that naturally involves singular functions, which are difficult to handle numerically. The aim of this part of our reasearch is to develop and analyze numerical methods for singular perturbation methods that are prone to high order numerical approximation, and robust with respect to the small parameter characterizing the singular perturbation.

## 4.4. Data analysis in astrophysics

We focus on computationally intensive numerical algorithms arising in the data analysis of current and forthcoming Cosmic Microwave Background (CMB) experiments in astrophysics. This application is studied in collaboration with researchers from University Paris Diderot, and the objective is to make available the algorithms to the astrophysics community, so that they can be used in large experiments.

In CMB data analysis, astrophysicists produce and analyze multi-frequency 2D images of the universe when it was 5% of its current age. The new generation of the CMB experiments observes the sky with thousands of detectors over many years, producing overwhelmingly large and complex data sets, which nearly double every year therefore following Moore's Law. Planck (http://planck.esa.int/) is a keystone satellite mission which has been developed under auspices of the European Space Agency (ESA). Planck has been surveying the sky since 2010, produces terabytes of data and requires 100 Petaflops per image analysis of the universe. It is predicted that future experiments will collect half petabyte of data, and will require 100 Exaflops per analysis as early as in 2020. This shows that data analysis in this area, as many other applications, will keep pushing the limit of available supercomputing power for the years to come.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

Laura Grigori was awarded with E. Cances, Y. Maday, and J.-P. Piquemal an ERC Syngergy Grant for the Extreme-scale Mathematically-based Computational Chemistry project (EMC2), 2018. A description of the project can be found here.

# 6. New Software and Platforms

## 6.1. FreeFem++

*FeeFrem++*
SCIENTIFIC DESCRIPTION: FreeFem++ is a partial differential equation solver. It has its own language. freefem scripts can solve multiphysics non linear systems in 2D and 3D.

Problems involving PDE (2d, 3d) from several branches of physics such as fluid-structure interactions require interpolations of data on several meshes and their manipulation within one program. FreeFem++ includes a fast 2dˆ-tree-based interpolation algorithm and a language for the manipulation of data on multiple meshes (as a follow up of bamg (now a part of FreeFem++ ).

FreeFem++ is written in C++ and the FreeFem++ language is a C++ idiom. It runs on Macs, Windows, Unix machines. FreeFem++ replaces the older freefem and freefem+.

FUNCTIONAL DESCRIPTION: FreeFem++ is a PDE (partial differential equation) solver based on a flexible language that allows a large number of problems to be expressed (elasticity, fluids, etc) with different finite element approximations on different meshes.

- Partner: UPMC
- Contact: Frederic Hecht
- URL: http://www.freefem.org/ff++/

## 6.2. HPDDM

SCIENTIFIC DESCRIPTION: HPDDM is an efficient implementation of various domain decomposition methods (DDM) such as one- and two-level Restricted Additive Schwarz methods, the Finite Element Tearing and Interconnecting (FETI) method, and the Balancing Domain Decomposition (BDD) method. This code has been proven to be efficient for solving various elliptic problems such as scalar diffusion equations, the system of linear elasticity, but also frequency domain problems like the Helmholtz equation. A comparison with modern multigrid methods can be found in the thesis of Pierre Jolivet.
FUNCTIONAL DESCRIPTION: HPDDM is an efficient implementation of various domain decomposition methods (DDM) such as one- and two-level Restricted Additive Schwarz methods, the Finite Element Tearing and Interconnecting (FETI) method, and the Balancing Domain Decomposition (BDD) method.

- Participants: Frédéric Nataf and Pierre Jolivet
- Contact: Pierre Jolivet
- URL: https://github.com/hpddm

## 6.3. LORASC

*LORASC preconditioner*
KEYWORD: Preconditioner

- Participants: Laura Grigori and Rémi Lacroix
- Contact: Laura Grigori

## 6.4. Platforms

### *6.4.1. HTOOL*

KEYWORD: Hierarchical Matrices
FUNCTIONAL DESCRIPTION: HTOOL is a C++ header-only library implementing compression techniques (e.g. Adaptive Cross Approximation) using hierarchical matrices. The library uses MPI and OpenMP for parallelism, and is interfaced with HPDDM for the solution of linear systems.

- Partners: CNRS - UPMC - ANR NonlocalDD
- Contact: Pierre Marchand
- URL: https://github.com/PierreMarchand20/htool

### *6.4.2. BemTool*

KEYWORD: Boundary Element Method
FUNCTIONAL DESCRIPTION: BemTool is a C++ header-only library implementing the boundary element method for the discretisation of the Laplace, Helmholtz and Maxwell equations, in 2D and 3D. Its main purpose is the assembly of classic boundary element matrices, which can be compressed and inverted through its interface with HTOOL.

- Partners: UPMC - ANR NonlocalDD
- Contact: Xavier Claeys
- URL: https://github.com/xclaeys/BemTool

### *6.4.3. Geneo4PETSc*

KEYWORD: Domain decomposition method
FUNCTIONAL DESCRIPTION: Implementation of the GenEO preconditioner with PETSc and SLEPc.

- Partners: CNRS - UPMC - European project NLAFET
- Contact: Frédéric Nataf
- URL: https://github.com/geneo4PETSc/geneo4PETSc

### *6.4.4. ffddm*

KEYWORD: Domain decomposition method
FUNCTIONAL DESCRIPTION: In the acronym ffddm, ff stands for FreeFem++ and ddm for domain decomposition methods. The idea behind ffddm is to simplify the use of parallel solvers in FreeFem++: distributed direct methods and domain decomposition methods.

- Partners: CNRS - UPMC
- Contact: Pierre-Henri Tournier and Frédéric Nataf
- URL: https://doc.freefem.org/documentation/ffddm/ffddm

### *6.4.5. preAlps*

KEYWORD: Preconditioned enlarged Krylov subspace methos
FUNCTIONAL DESCRIPTION: Contains enlarged Conjugate Gradient Krylov subspace method and Lorasc preconditioner.

- Partners: Inria
- Contact: Simplice Donfack, Laura Grigori, Olivier Tissot
- URL: https://github.com/NLAFET/preAlps

### *6.4.6. FreeFem++ v4*

KEYWORD: New version of Freefem++, with new sparce matrix kernel, and with surface finite element.
FUNCTIONAL DESCRIPTION:

- Partners: UPMC - Inria
- Contact: Frederic Hecht
- URL: https://github.com/FreeFem/FreeFem-sources/tree/v4

# 7. New Results

## 7.1. First kind Galerkin boundary element method for the Hodge-Laplacian in three dimensions

Boundary value problems for the Euclidean Hodge-Laplacian in three dimension $-\Delta_{HL} = \mathbf{curlcurl} - \mathbf{grad}\mathrm{div}$ lead to variational formulations set in subspaces of $\mathbf{H}(\mathbf{curl},\Omega) \cap \mathbf{H}(\mathrm{div},\Omega)$, $\Omega \subset \mathbb{R}^3$ a bounded Lipschitz domain. Via a representation formula and Calderón identities we derive corresponding first-kind boundary integral equations set in trace spaces of $H^1(\Omega)$, $\mathbf{H}(\mathbf{curl},\Omega)$, and $\mathbf{H}(\mathrm{div},\Omega)$. They give rise to saddle-point variational formulations and feature kernels whose dimensions are linked to fundamental topological invariants of $\Omega$.

Kernels of the same dimensions also arise for the linear systems generated by low-order conforming Galerkin boundary element (BE) discretization. On their complements, we can prove stability of the discretized problems, nevertheless. We prove that discretization does not affect the dimensions of the kernels and also illustrate this fact by numerical tests.

## 7.2. Boundary integral multi-trace formulations and Optimised Schwarz Methods

In the present contribution, we consider Helmholtz equation with material coefficients being constant in each subdomain of a geometric partition of the propagation medium (discarding the presence of junctions), and we are interested in the numerical solution of such a problem by means of local multi-trace boundary integral formulations (local-MTF). For a one dimensional problem and configurations with two subdomains, it has been recently established that applying a Jacobi iterative solver to local-MTF is exactly equivalent to an Optimised Schwarz Method (OSM) with a non-local impendance. In the present contribution, we show that this correspondance still holds in the case where the subdomain partition involves an arbitrary number of subdomains. From this, we deduce that the depth of the adjacency graph of the subdomain partition plays a critical role in the convergence of linear solvers applied to local-MTF: we prove it for the case of homogeneous propagation medium and show, through numerical evidences, that this conclusion still holds for heterogeneous media. Our study also shows that, considering variants of local-MTF involving a relaxation parameter, there is a fixed value of this relaxation parameter that systematically leads to optimal speed of convergence for linear solvers.

## 7.3. Poroelasticity

In [38], we design and study a fully coupled numerical scheme for the poroelasticity problem modelled through Biot's equations. The classical way to numerically solve this system is to use a finite element method for the mechanical equilibrium equation and a finite volume method for the fluid mass conservation equation. However, to capture specific properties of underground media such as heterogeneities, discontinuities and faults, meshing procedures commonly lead to badly shaped cells for finite element based modelling. Consequently, we investigate the use of the recent virtual element method which appears as a potential discretization method for the mechanical part and could therefore allow the use of a unique mesh for the both mechanical and fluid flow modelling. Starting from a first insight into virtual element method applied to the elastic problem in the context of geomechanical simulations, we apply in addition a finite volume method to take care of the fluid conservation equation. We focus on the first order virtual element method and the two point flux approximation for the finite volume part. A mathematical analysis of this original coupled scheme is provided, including existence and uniqueness results and a priori estimates. The method is then illustrated by some computations on two or three dimensional grids inspired by realistic application cases.

## 7.4. Hybrid discontinuous Galerkin discretisation and domain decomposition preconditioners for the Stokes problem

Solving the Stokes equation by an optimal domain decomposition method derived algebraically involves the use of nonstandard interface conditions whose discretisation is not trivial. For this reason the use of approximation methods such as hybrid discontinuous Galerkin appears as an appropriate strategy: on the one hand they provide the best compromise in terms of the number of degrees of freedom in between standard continuous and discontinuous Galerkin methods, and on the other hand the degrees of freedom used in the nonstandard interface conditions are naturally defined at the boundary between elements. In this paper, we introduce the coupling between a well chosen discretisation method (hybrid discontinuous Galerkin) and a novel and efficient domain decomposition method to solve the Stokes system. We present the detailed analysis of the hybrid discontinuous Galerkin method for the Stokes problem with non standard boundary conditions. This analysis is supported by numerical evidence. In addition, the advantage of the new preconditioners over more classical choices is also supported by numerical experiments. The full paper [18] is available at https://hal.archives-ouvertes.fr/hal-01967577

## 7.5. A class of efficient locally constructed preconditioners based on coarse spaces

In [14] we present a class of robust and fully algebraic two-level preconditioners for SPD matrices. We introduce the notion of algebraic local SPSD splitting of an SPD matrix and we give a characterization of this splitting. This splitting leads to construct algebraically and locally a class of efficient coarse spaces which bound the spectral condition number of the preconditioned matrix by a number defined a priori. We also introduce the notion of filtering subspace. This concept helps compare the dimension minimality of coarse spaces. Some PDEs-dependant preconditioners correspond to a special case. The examples of the algebraic coarse spaces in this paper are not practical due to expensive construction. We propose a heuristic approximation that is not costly. Numerical experiments illustrate the efficiency of the proposed method.

## 7.6. Enlarged Krylov methods for reducing communication

Krylov methods are widely used for solving large sparse linear systems of equations. On distributed architectures, their performance is limited by the communication needed at each iteration of the algorithm. In [34], we study the use of so-called enlarged Krylov subspaces for reducing the number of iterations, and therefore the overall communication, of Krylov methods. In particular, we consider a reformulation of the Conjugate Gradient method using these enlarged Krylov subspaces: the enlarged Conjugate Gradient method. We present the parallel design of two variants of the enlarged Conjugate Gradient method as well as their corresponding dynamic versions where the number of search directions is dynamically reduced during the iterations. For a linear elasticity problem with heterogeneous coefficients using a block Jacobi preconditioner, we show that this implementation scales up to 16,384 cores, and is up to 6,9 times faster than the PETSc implementation of PCG.

In [15] we propose a variant of the GMRES method for solving linear systems of equations with one or multiple right-hand sides. Our method is based on the idea of the enlarged Krylov subspace to reduce communication. It can be interpreted as a block GMRES method. Hence, we are interested in detecting inexact breakdowns. We introduce a strategy to perform the test of detection. Furthermore, we propose an eigenvalues deflation technique aiming to have two benefits. The first advantage is to avoid the plateau of convergence after the end of a cycle in the restarted version. The second is to have a very fast convergence when solving the same system with different right-hand sides, each given at a different time (useful in the context of CPR preconditioner). With the same memory cost, we obtain a saving of up to $50\%$ in the number of iterations to reach convergence with respect to the original method.

## 7.7. Recycling Krylov subspaces and reducing deflation subspaces for solving a sequence of linear systems

In [32] we present deflation strategies related to recycling Krylov subspace methods for solving one or a sequence of linear systems of equations. Besides well-known strategies of deflation, Ritz and harmonic Ritz based deflation, we introduce an SVD-based deflation technique. We consider the recycling in two contexts, recycling the Krylov subspace between the cycles of restarts and recycling a deflation subspace when the matrix changes in a sequence of linear systems. Numerical experiments on real-life reservoir simulations demonstrate the impact of our proposed strategy.

## 7.8. Solving linear equations with messenger-field and conjugate gradient techniques: an application to CMB data analysis

In [26] we discuss linear system solvers invoking a messenger-field and compare them with (preconditioned) conjugate gradients approaches. We show that the messenger-field techniques correspond to fixed point iterations of an appropriately preconditioned initial system of linear equations. We then argue that a conjugate gradient solver applied to the same preconditioned system, or equivalently a preconditioned conjugate gradient solver using the same preconditioner and applied to the original system, will in general ensure at least a comparable and typically better performance in terms of the number of iterations to convergence and time-to-solution. We illustrate our conclusions on two common examples drawn from the Cosmic Microwave Background data analysis: Wiener filtering and map-making. In addition, and contrary to the standard lore

in the CMB field, we show that the performance of the preconditioned conjugate gradient solver can depend importantly on the starting vector. This observation seems of particular importance in the cases of map-making of high signal-to-noise sky maps and therefore should be of relevance for the next generation of CMB experiments.

## 7.9. Low rank approximation of a sparse matrix based on LU factorization with column and row tournament pivoting

In [23] we present an algorithm for computing a low rank approximation of a sparse matrix based on a truncated LU factorization with column and row permutations. We present various approaches for determining the column and row permutations that show a trade-off between speed versus deterministic/probabilistic accuracy. We show that if the permutations are chosen by using tournament pivoting based on QR factorization, then the obtained truncated LU factorization with column/row tournament pivoting, LU_CRTP, satisfies bounds on the singular values which have similarities with the ones obtained by a communication avoiding rank revealing QR factorization. Experiments on challenging matrices show that LU_CRTP provides a good low rank approximation of the input matrix and it is less expensive than the rank revealing QR factorization in terms of computational and memory usage costs, while also minimizing the communication cost. We also compare the computational complexity of our algorithm with randomized algorithms and show that for sparse matrices and high enough but still modest accuracies, our approach is faster.

## 7.10. ALORA: affine low-rank approximations

In [17] we introduce the concept of affine low-rank approximation for an $m \times n$ matrix, consisting in fitting its columns into an affine subspace of dimension at most $k << \min(m, n)$. We show that the optimal affine approximation can be obtained by applying an orthogonal projection to the matrix before constructing its best approximation. Moreover, we present the algorithm ALORA that constructs an affine approximation by slightly modifying the application of any low-rank approximation method. We focus on approximations created with the classical QRCP and subspace iteration algorithms. For the former, we present a detailed analysis of the existing pivoting techniques and furthermore, we provide a bound for the error when an arbitrary pivoting technique is used. For the case of subspace iteration, we prove a result on the convergence of singular vectors, showing a bound that is in agreement with the one for convergence of singular values proved recently. Finally, we present numerical experiences using challenging matrices taken from different fields, showing good performance and validating the theoretical framework.

## 7.11. Linear-time CUR approximation of BEM matrices

In [33] we propose linear-time CUR approximation algorithms for admissible matrices obtained from the hierarchical form of Boundary Element matrices. We propose a new approach called geometric sampling to obtain indices of most significant rows and columns using information from the domains where the problem is posed. Our strategy is tailored to Boundary Element Methods (BEM) since it uses directly and explicitly the cluster tree containing information from the problem geometry. Our CUR algorithm has precision comparable with low-rank approximations created with the truncated QR factorization with column pivoting (QRCP) and the Adaptive Cross Approximation (ACA) with full pivoting, which are quadratic-cost methods. When compared to the well-known linear-time algorithm ACA with partial pivoting, we show that our algorithm improves, in general, the convergence error and overcomes some cases where ACA fails. We provide a general relative error bound for CUR approximations created with geometrical sampling. Finally, we evaluate the performance of our algorithms on traditional BEM problems defined over different geometries.

## 7.12. Fractional decomposition of matrices and parallel computing

In [40] we are interested in the design of parallel numerical schemes for linear systems. We give an effective solution to this problem in the following case: the matrix $A$ of the linear system is the product of $p$ nonsingular matrices $A_i^m$ with specific shape: $A_i = I - h_i X$ for a fixed matrix $X$ and real numbers $h_i$. Although having

the special form, these matrices $A_i$ arise frequently in the discretization of evolutionary Partial Differential Equations. The idea is to express $A^{-1}$ as a linear combination of elementary matrices $A_i^{-k}$. Hence the solution of the linear system with matrix $A$ is a linear combination of the solutions of linear systems with matrices $A_i^k$. These systems are solved simultaneously on different processors.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

- Contract with Total, February 2015 - August 2018, that funds the PhD thesis of Hussam Al Daas on enlarged Krylov subspace methods for oil reservoir and seismic imaging applications. Supervisor L. Grigori.
- Contract with IFPen, February 2016 - April 2019, that funds the Phd thesis of Zakariae Jorti on adaptive preconditioners using a posteriori error estimators. Supervisor L. Grigori.
- Contract with IFPen, October 2016 - October 2019, that funds the Phd thesis of Julien Coulet on the virtual element method (VEM). Supervisor F. Nataf and V. Girault.
- Contract with Total, February - September 2018, that funded an internship on Helmholtz domain decomposition solvers for multiple right hand sides. Supervisor F. Nataf.

# 9. Partnerships and Cooperations

## 9.1. Regional Initiatives

GIS, Géosciences franciliennes: scientific collaboration network between ten public institutions from the Paris (Ile-de-France) region, focused on natural resources and environment. The project-team Alpines is a member.

## 9.2. National Initiatives

### *9.2.1. ANR*

#### *9.2.1.1. B3DCMB*

ANR Decembre 2017 - Novembre 2021 This project is in the area of data analysis of cosmological data sets as collected by contemporary and forthcoming observatories. This is one of the most dynamic areas of modern cosmology. Our special target are data sets of Cosmic Microwave Background (CMB) anisotropies, measurements of which have been one of the most fruitful of cosmological probes. CMB photons are remnants of the very early evolution of the Universe and carry information about its physical state at the time when the Universe was much younger, hotter and denser, and simpler to model mathematically. The CMB has been, and continue to be, a unique source of information for modern cosmology and fundamental physics. The main objective of this project is to empower the CMB data analysis with novel high performance tools and algorithms superior to those available today and which are capable of overcoming the existing performance gap. Partners: AstroParticules et Cosmologie Paris 7 (PI R. Stompor), ENSAE Paris Saclay.

#### *9.2.1.2. ANR Cine-Para*

October 2015 - September 2019, Laura Grigori is Principal Coordinator for Inria Paris. Funding for Inria Paris is 145 Keuros. The funding for Inria is to combine Krylov subspace methods with parallel in time methods. Partners: University Pierre and Marie Curie, J. L. Lions Laboratory (PI Y. Maday), CEA, Paris Dauphine University, Paris 13 University.

#### *9.2.1.3. Non-local DD*

ANR appel à projet générique October 2015 - September 2020

This project in scientific computing aims at developing new domain decomposition methods for massively parallel simulation of electromagnetic waves in harmonic regime. The specificity of the approach that we propose lies in the use of integral operators not only for solutions local to each subdomain, but for coupling subdomains as well. The novelty of this project consists, on the one hand, in exploiting multi-trace formalism for domain decomposition and, on the other hand, considering optimized Schwarz methods relying on Robin type transmission conditions involving quasi-local integral operators.

*9.2.1.4. Soil$\mu$-3D*

ANR appel à projet générique October 2015 - September 2020

In spite of decades of work on the modeling of greenhouse gas emission such as CO2 and N2O and on the feedback effects of temperature and water content on soil carbon and nitrogen transformations, there is no agreement on how these processes should be described, and models are widely conflicting in their predictions. Models need improvements to obtain more accurate and robust predictions, especially in the context of climate change, which will affect soil moisture regime.

The goal of this new project is now to go further using the models developed in MEPSOM to upscale heterogeneities identified at the scale of microbial habitats and to produce macroscopic factors for biogeochemical models running at the field scale.

To achieve this aim, it will be necessary to work at different scales: the micro-scale of pores ($\mu$m) where the microbial habitats are localized, the meso-scale of cores at which laboratory measurements on CO2 and N2O fluxes can be performed, and the macro-scale of the soil profile at which outputs are expected to predict greenhouse gas emission. The aims of the project are to (i) develop new descriptors of the micro-scale 3D soil architecture that explain the fluxes measured at the macro-scale, (ii) Improve the performance of our 3D pore scale models to simulate both micro-and meso- scales at the same time. Upscaling methods like "homogeneization" would help to simulate centimeter samples which cannot be achieved now. The reduction of the computational time used to solve the diffusion equations and increase the number of computational units, (iii) develop new macro-functions describing the soil micro-heterogeneity and integrate these features into the field scale models.

## 9.3. European Initiatives

### *9.3.1. FP7 & H2020 Projects*

*9.3.1.1. NLAFET (197)*

Title: Parallel Numerical Linear Algebra for Future Extreme-Scale Systems

Programm: H2020

Duration: November 2015 - April 2019

Coordinator: UMEÅ Universitet

Partners:

Science and Technology Facilities Council (United Kingdom)

Computer Science Department, UmeåUniversitet (Sweden)

Mathematics Department, The University of Manchester (United Kingdom)

Inria, Alpines group

Inria contact: Laura Grigori

The NLAFET proposal is a direct response to the demands for new mathematical and algorithmic approaches for applications on extreme scale systems, as identified in the FETHPC work programme and call. This project will enable a radical improvement in the performance and scalability of a wide range of real-world applications relying on linear algebra software, by developing novel architecture-aware algorithms and software libraries, and the supporting runtime capabilities to achieve scalable performance and resilience on heterogeneous architectures. The focus is on a

critical set of fundamental linear algebra operations including direct and iterative solvers for dense and sparse linear systems of equations and eigenvalue problems. Achieving this requires a co-design effort due to the characteristics and overwhelming complexity and immense scale of such systems. Recognized experts in algorithm design and theory, parallelism, and auto-tuning will work together to explore and negotiate the necessary tradeoffs. The main research objectives are: (i) development of novel algorithms that expose as much parallelism as possible, exploit heterogeneity, avoid communication bottlenecks, respond to escalating fault rates, and help meet emerging power constraints; (ii) exploration of advanced scheduling strategies and runtime systems focusing on the extreme scale and strong scalability in multi/many-core and hybrid environments; (iii) design and evaluation of novel strategies and software support for both offline and online auto-tuning. The validation and dissemination of results will be done by integrating new software solutions into challenging scientific applications in materials science, power systems, study of energy solutions, and data analysis in astrophysics. The deliverables also include a sustainable set of methods and tools for cross-cutting issues such as scheduling, auto-tuning, and algorithm-based fault tolerance packaged into open-source library modules.

# 9.4. International Initiatives

## *9.4.1. Inria International Partners*

*9.4.1.1. Informal International Partners*

- J. Demmel, UC Berkeley, USA
- R. Hipmair, ETH Zurich
- M. Grote, Université de Bâle, Suisse
- F. Assous, Israel

# 9.5. International Research Visitors

## *9.5.1. Visits of International Scientists*

- Visit to Xavier Claeys of Jan Zapletal from IT4Innovation of University of Ostrava, Czech Republic from 4th to 30th of March 2018. The main topic of the visit was discussions around HPC implementation of multi-trace formulations in the BEM code of IT4Innovation.
- Visit to Laura Grigori of Agnieszka Miedlar, University of Kansas, from Jun 2018 until Jul 2018.
- Visit to Laura Grigori of Qiang Niu, Xi'an Jiaotong Liverpool University, from May 2018 until Jul 2018.
- Visit to Frédéric Nataf of Lawrence Mitchell from University of Durham (UK) from December 17th to 22nd. The main topic of the visit was to finalize the interface of the finite element software Firedrake to our library geneo4PETSc.
- Visit to Frédéric Hecht of T. Chacon of Differential equations and numerical analysis at University of Seville Rectorate form April 23th to May 4th.
- Visit to Frédéric Hecht of P. Degond of Department of Mathematics at Imperial College London form Juin 6th to 10th.

*9.5.1.1. Internships*

- Visit to Xavier Claeys of Michal Kravchenko from IT4Innovation of University of Ostrava, Czech Republic from 1st of October to 28th of December 2018. The main subject of the visit was effective implementation of multi-trace formulations in the BEM code of IT4Innovation.

## *9.5.2. Visits to International Teams*

*9.5.2.1. Research Stays Abroad*

- Visit of Xavier Claeys to Ralf Hiptmair at ETH Zuerich from the 19th of August to 25th of August 2018. The main subject of the visit was discussion on boundary integral equations adapted to low frequency electromagnetics.
- Visit of Xavier Claeys to Paul Escapil-Inchauspe at Pontificia Universidad Catholica at Santiago Chile for further collaboration around analysis of local multi-trace formulation for electromagnetics.
- Visit of Laura Grigori to the group of Professor J. Demmel, UC Berkeley, for 6 weeks in July and August 2018.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

- Xavier Claeys was co-chair of the "Symposium of the International Association for Boundary Element Methods" in June 26-28 2018, an international conference that took place in Jussieu campus of Sorbonne Université and hosted 140 participants.
- Frederic Hecht organized the 10th FreeFem++ days (December 12-14, 2018, Paris), https://freefem.org/ff-days/

### *10.1.2. Journal*

*10.1.2.1. Member of the Editorial Boards*

- Laura Grigori, March 2014 – current. Member of the editorial board for the SIAM book series Software, Environments and Tools. See http://bookstore.siam.org/software-environments-and-tools/.
- Laura Grigori, January 2016 – current. Associate Editor, SIAM Journal on Scientific Computing.
- Laura Grigori, January 2017 – current. Associate Editor, SIAM Journal on Matrix Analysis and Applications.
- Laura Grigori, January 2016 – current. Editorial board, Numerical linear algebra with applications Journal, Wiley.
- Frédéric Nataf, January 2015 – current, Editorial board, Journal of Numerical Mathematics, de Gruyter.

### *10.1.3. Invited Talks*

- Xavier Claeys was invited speaker at the second national congress of the Société Mahématique de France (SMF) in June 2018.
- Laura Grigori was
  - Keynote speaker, International Symposium on Computational Science at Scale, September 2018, Erlangen-Nurnberg Germany.
  - Invited plenary speaker, SIAM Conference on Applied Linear Algebra, Hong Kong May 2018.
- Frédéric Nataf was invited speaker at
  - Workshop for Robert Scheichl's farewell, Bath University, November 2018.
  - NUMACH 2018: Numerical Methods for Challenging Problems, Mulhouse (France) July 2018.
  - 10th International Workshop on Parallel Matrix Algorithms and Applications (PMAA'18) in ETH Zurich (Switzerland) June 2018.

- Frédéric Hecht was invited speaker at
  - XVIII Spanish-French school Jacques-Louis Lions about numerical simulation in physics and engineering, Las Palmas de Gran Canaria, 25-29 June 2018.

### *10.1.4. Leadership within the Scientific Community*

- Laura Grigori, member elected of SIAM Council, January 2018 - December 2020, the committee supervising the scientific activities of SIAM. Nominated by a Committee and elected by the members of SIAM.
- Laura Grigori, member of the PRACE (Partnership for Advanced Computing in Europe) Scientific Steering Committee, September 2016 - current.
- Laura Grigori and Frédéric Hecht are coordinators of the High Performance in Scientific Computing Major of second year of Mathematics and Applications Master, Sorbonne University.

### *10.1.5. Scientific Expertise*

- Laura Grigori: November 2015 - current, expert to the Scientific Commission of IFPEN (French Petroleum Institute). Evaluation of research programs, PhD theses, work representing a total of 5 days per year.

### *10.1.6. Research Administration*

- Laura Grigori is vice-president of the committee CE46 of ANR, September 2017 - July 2018.
- Frédéric Nataf is president of the committee CE40 of ANR, September 2017 - current.

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master 2: Laura Grigori, Course on *High performance computing, large scale linear algebra, and numerical stability* (*Calcul haute performance, algorithmes parallèles d'algèbre linéaire à grande echelle, stabilité numérique* in french), https://who.rocq.inria.fr/Laura.Grigori/TeachingDocs/UPMC_Master2/Spr2018.html, Master 2nd year, Mathematics & Applications, UPMC, 24 hours of lectures per year.

Master 2: Laura Grigori, Winter 2018, Participation in the course on High Performance Computing given at UPMC, Computer Science, intervention for 8 hours per year.

Master 2: Laura Grigori, Course on *High performance computing for numerical methods and data analysis*, https://who.rocq.inria.fr/Laura.Grigori/TeachingDocs/UPMC_Master2/HPC_MN_DA.html, Master 2nd year, Mathematics & Applications, UPMC, 24 hours per year.

Master 1: Xavier Claeys, supervision of a student project for a group of 4 students in the curriculum Polytech, 40hrs, UPMC.

Master 1: Xavier Claeys, Initiation to C++, 36 hrs of programming tutorials in C++, UPMC.

Master 1: Xavier Claeys, Computational Linear Algebra, 32 hrs of lectures, UPMC.

Master 1: Xavier Claeys, Approximation of EDPs, 24 hrs of programming tutorials in Python, UPMC.

Master 2: Frédéric Nataf, Course on Domain Decomposition Methods, UPMC

Master 1: Frédéric Hecht, Initiation au C++, 24hrs, UPMC, France

Master 2: Frédéric Hecht, Des EDP à leur résolution par la méthode des éléments finis (MEF), 36hrs, M2, UPMC, France

Master 2: Frédéric Hecht, Numerical methods for fluid mechanics, 10hrs, UPMC, France

Master 2: Frédéric Hecht, Calcul scientifique 3 / projet industriel FreeFem++, 28hrs, M2, UPMC, France

Master 2: Frédéric Hecht, Ingénierie 1 / Logiciel pour la simulation (FreeFem++), 21hrs, UPMC, France

Master 2: Frédéric Hecht, Ingénierie 2 / Projet collaboratif, 21hrs, UPMC, France

### *10.2.2. Supervision*

PhD: Alan Ayala, Complexity reduction methods applied to the rapid solution to multi-trace boundary integral formulations, Sorbonne Université, November 2018 (funded by NLAFET H2020 project), co-advisors Xavier Claeys and Laura Grigori.

PhD: Hussam Al Daas, Solving linear systems arising from reservoirs modelling, Sorbonne Université, December 2018, (funded by contract with Total), advisor Laura Grigori.

PhD in progress : Sebastien Cayrols, since October 2013 (funded by Maison de la simulation), adivsor Laura Grigori.

PhD in progress: Olivier Tissot, since October 2015 (funded by NLAFET H2020 project), advisor Laura Grigori.

PhD in progress: Rim El Dbaissy, since November 2015 (funded by Univ. St Joseph, Liban), advisors Tony Sayah, Frédéric Hecht.

PhD in progress: Pierre Marchand, since October 2016 (funded by ANR NonLocalDD project), advisors Xavier Claeys and Frédéric Nataf.

PhD in progress: Zakariae Jorti, since February 2016 (funded by IFPen), advisor Laura Grigori.

PhD in progress: Igor Chollet, since October 2017 (funded by ICSD), advisors Xavier Claeys, Pierre Fortin, Laura Grigori.

PhD in progress: Thanh Van Nguyen, since November 2017 (funded by ANR CinePara), advisor Laura Grigori.

### *10.2.3. Juries*

- Xavier Claeys was examiner at the PhD defense of Wen Xu on the 17th of July 2018 at École Centrale Supélec. Title of the thesis: "Relevant numerical methods for mesoscale wave propagation in heterogeneous media".
- Laura Grigori was examiner of the Phd defense of Gilles Moreau, ENS Lyon, December 2018.
- Laura Grigori was president of the HDR habilitation defense of Pierre Fortin, Sorbonne University, July 2018.
- Laura Grigori was examiner of the Phd defense of Amanda Bienz, June 2018, University of Illinois at Urbana Champaign.
- Frédéric Nataf was examiner at the Phd defense of Louis Viot, 2018, ENS Cachan
- Frédéric Nataf was president of the PhD defense of H. Al Daas, 2018, UPMC
- Frédéric Hecht was referee of the HDR habilitation defense of S. Glockner, 2018, I2M, Bordeaux
- Frédéric Hecht was referee of the PhD defense of G. Dollé, 2018, Univ. Strasbourg
- Frédéric Hecht was examiner at the Phd defense of G. Morel, 2018, Sorbonne University

## 10.3. Popularization

### *10.3.1. Internal or external Inria responsibilities*

Laura Grigori is vice-president of the Evaluation Commission of Inria, March 2018 - current.

# 11. Bibliography

## Major publications by the team in recent years

[1] X. CLAEYS.*Essential spectrum of local multi-trace boundary integral operators*, in "IMA J. Appl. Math.", 2016, vol. 81, n$^{o}$ 6, p. 961–983, https://doi-org.accesdistant.sorbonne-universite.fr/10.1093/imamat/hxw019

[2] X. CLAEYS, R. HIPTMAIR.*Integral equations for electromagnetic scattering at multi-screens*, in "Integral Equations Operator Theory", 2016, vol. 84, n$^{o}$ 1, p. 33–68, https://doi-org.accesdistant.sorbonne-universite.fr/10.1007/s00020-015-2242-5

[3] X. CLAEYS, R. HIPTMAIR, E. SPINDLER.*Second kind boundary integral equation for multi-subdomain diffusion problems*, in "Adv. Comput. Math.", 2017, vol. 43, n$^{o}$ 5, p. 1075–1101, https://doi-org.accesdistant.sorbonne-universite.fr/10.1007/s10444-017-9517-0

[4] J. W. DEMMEL, L. GRIGORI, M. HOEMMEN, J. LANGOU.*Communication-optimal parallel and sequential QR and LU factorizations*, in "SIAM Journal on Scientific Computing", 2012, n$^{o}$ 1, p. 206-239, short version of technical report UCB/EECS-2008-89 from 2008

[5] V. DOLEAN, P. JOLIVET, F. NATAF.*An Introduction to Domain Decomposition Methods: algorithms, theory and parallel implementation*, SIAM, 2015

[6] L. GRIGORI, J. DEMMEL, H. XIANG.*CALU: a communication optimal LU factorization algorithm*, in "SIAM Journal on Matrix Analysis and Applications", 2011, vol. 32, p. 1317-1350

[7] L. GRIGORI, S. MOUFAWAD, F. NATAF.*Enlarged Krylov Subspace Conjugate Gradient methods for Reducing Communication*, in "SIAM Journal on Matrix Analysis and Applications", 2016, vol. 37, n$^{o}$ 2, p. 744-773

[8] R. HAFERSSAS, P. JOLIVET, F. NATAF.*An Additive Schwarz Method Type Theory for Lions's Algorithm and a Symmetrized Optimized Restricted Additive Schwarz Method*, in "SIAM J. Sci. Comput.", 2017, vol. 39, n$^{o}$ 4, p. A1345–A1365, http://dx.doi.org/10.1137/16M1060066

[9] F. HECHT.*New development in FreeFem++*, in "J. Numer. Math.", 2012, vol. 20, n$^{o}$ 3-4, p. 251–265

[10] P. JOLIVET, F. NATAF.*HPDDM: High-Performance Unified framework for Domain Decomposition methods, MPI-C++ library*, 2014, https://github.com/hpddm/hpddm

[11] N. SPILLANE, V. DOLEAN, P. HAURET, F. NATAF, C. PECHSTEIN, R. SCHEICHL.*Abstract robust coarse spaces for systems of PDEs via generalized eigenproblems in the overlaps*, in "Numer. Math.", 2014, vol. 126, n$^{o}$ 4, p. 741–770, http://dx.doi.org/10.1007/s00211-013-0576-y

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[12] H. AL DAAS.*Solving linear systems arising from reservoirs modelling*, Inria Paris ; Sorbonne Université, UPMC University of Paris 6, Laboratoire Jacques-Louis Lions, December 2018, https://hal.inria.fr/tel-01984047

[13] A. A. OBREGÓN.*Complexity reduction methods applied to the rapid solution to multi-trace boundary integral formulations*, Sorbonne University , UPMC, November 2018, https://tel.archives-ouvertes.fr/tel-02004298

### Articles in International Peer-Reviewed Journal

[14] H. AL DAAS, L. GRIGORI.*A class of efficient locally constructed preconditioners based on coarse spaces*, in "SIAM Journal on Matrix Analysis and Applications", 2018, https://hal.inria.fr/hal-01963067

[15] H. AL DAAS, L. GRIGORI, P. HÉNON, P. RICOUX.*Enlarged GMRES for solving linear systems with one or multiple right-hand sides*, in "IMA Journal of Numerical Analysis", August 2018, https://hal.inria.fr/hal-01963032

[16] R. ALDBAISSY, F. HECHT, G. MANSOUR, T. SAYAH.*A full discretisation of the time-dependent Boussinesq (buoyancy) model with nonlinear viscosity*, in "Calcolo", December 2018, vol. 55, n$^{o}$ 4, https://hal.archives-ouvertes.fr/hal-01972178

[17] A. AYALA, X. CLAEYS, L. GRIGORI.*ALORA: Affine Low-Rank Approximations*, in "Journal of Scientific Computing", April 2018, https://hal.inria.fr/hal-01762882

[18] G. R. BARRENECHEA, V. DOLEAN, F. NATAF, P.-H. TOURNIER.*Hybrid discontinuous Galerkin discretisation and domain decomposition preconditioners for the Stokes problem*, in "Computational Methods in Applied Mathematics", March 2018, https://hal.archives-ouvertes.fr/hal-01967577

[19] M. BONAZZOLI, V. DOLEAN, F. HECHT, F. RAPETTI.*An example of explicit implementation strategy and preconditioning for the high order edge finite elements applied to the time-harmonic Maxwell's equations*, in "Computers and Mathematics with Applications", March 2018, vol. 75, n$^{o}$ 5, p. 1498 - 1514, https://arxiv.org/abs/1711.05629 [*DOI :* 10.1016/J.CAMWA.2017.11.013], https://hal.archives-ouvertes.fr/hal-01298938

[20] M. BONAZZOLI, V. DOLEAN, F. RAPETTI, P.-H. TOURNIER.*Parallel preconditioners for high order discretizations arising from full system modeling for brain microwave imaging*, in "International Journal of Numerical Modelling: Electronic Networks, Devices and Fields", 2018, vol. 31, n$^{o}$ 2 [*DOI :* 10.1002/JNM.2229], https://hal.archives-ouvertes.fr/hal-01328197

[21] T. CHACON REBOLLO, M. GÓMEZ MARMOL, F. HECHT, S. RUBINO, I. SÁNCHEZ MUÑOZ.*A High-Order Local Projection Stabilization Method for Natural Convection Problems*, in "Journal of Scientific Computing", February 2018, vol. 74, n$^{o}$ 2, p. 667-692 [*DOI :* 10.1007/S10915-017-0469-9], https://hal.archives-ouvertes.fr/hal-01972136

[22] L. CHESNEL, X. CLAEYS, S. A. NAZAROV.*Oscillating behaviour of the spectrum for a plasmonic problem in a domain with a rounded corner*, in "ESAIM: Mathematical Modelling and Numerical Analysis", September 2018 [*DOI :* 10.1051/M2AN/2016080], https://hal.archives-ouvertes.fr/hal-01240977

[23] L. GRIGORI, S. CAYROLS, J. W. DEMMEL.*Low Rank Approximation of a Sparse Matrix Based on LU Factorization with Column and Row Tournament Pivoting*, in "SIAM Journal on Scientific Computing", January 2018, vol. 40, n$^{o}$ 2, p. C181-C209, https://hal.archives-ouvertes.fr/hal-01967901

[24] F. HECHT, T. JANGVELADZE, Z. KIGURADZE, O. PIRONNEAU.*Finite difference scheme for one system of nonlinear partial integro-differential equations*, in "Applied Mathematics and Computation", July 2018, vol. 328, p. 287-300, https://hal.archives-ouvertes.fr/hal-01972158

[25] F. HECHT, I. NAJI, Z. MGHAZLI, J. ROBERTS.*A Residual a Posteriori Error Estimators for a Model for Flow in Porous Media with Fractures*, in "Journal of Scientific Computing", November 2018, p. 1–34 [*DOI :* 10.1007/s10915-018-0875-7], https://hal.sorbonne-universite.fr/hal-01972116

[26] J. PAPEZ, L. GRIGORI, R. STOMPOR.*Solving linear equations with messenger-field and conjugate gradient techniques: An application to CMB data analysis*, in "Astronomy and Astrophysics", November 2018, vol. 620, A59 [*DOI :* 10.1051/0004-6361/201832987], https://hal.archives-ouvertes.fr/hal-01731325

[27] M. ČERMÁK, F. HECHT, Z. TANG, M. VOHRALÍK.*Adaptive inexact iterative algorithms based on polynomial-degree-robust a posteriori estimates for the Stokes problem*, in "Numerische Mathematik", February 2018, vol. 138, n$^{o}$ 4, p. 1027-1065 [*DOI :* 10.1007/s00211-017-0925-3], https://hal.inria.fr/hal-01097662

### International Conferences with Proceedings

[28] G. BALLARD, J. W. DEMMEL, L. GRIGORI, M. JACQUELIN, N. KNIGHT.*A 3D Parallel Algorithm for QR Decomposition*, in "SPAA '18 - 30th ACM Symposium on Parallelism in Algorithms and Architectures", Vienna, Austria, ACM, July 2018, https://hal.inria.fr/hal-01968376

[29] V. L. COLI, P.-H. TOURNIER, V. DOLEAN, I. EL KANFOUD, C. PICHOT, C. MIGLIACCIO, L. BLANC-FÉRAUD.*Detection of Brain Strokes Using Microwave Tomography*, in "International Symposium on Antennas and Propagation and USNC-URSI Radio Science Meeting", Boston, United States, IEEE, July 2018, p. 223-224 [*DOI :* 10.1109/APUSNCURSINRSM.2018.8609404], https://hal.archives-ouvertes.fr/hal-01824526

### Scientific Books (or Scientific Book chapters)

[30] D. BOFFI, F. HECHT, O. PIRONNEAU.*Distributed Lagrange Multiplier for Fluid-Structure Interactions*, in "Numerical Methods for PDEs: State of the Art Techniques", Springer, October 2018, p. 129-145 [*DOI :* 10.1007/978-3-319-94676-4_5], https://hal.archives-ouvertes.fr/hal-01972151

### Research Reports

[31] H. AL DAAS, L. GRIGORI.*A class of efficient locally constructed preconditioners based on coarse spaces*, Inria – Centre Paris-Rocquencourt ; Laboratoire Jacques-Louis Lions, UPMC, Paris, June 2018, n$^{o}$ RR-9184, https://hal.inria.fr/hal-01816513

[32] H. AL DAAS, L. GRIGORI, P. HÉNON, P. RICOUX.*Recycling Krylov subspaces and reducing deflation subspaces for solving sequence of linear systems*, Inria Paris, October 2018, n$^{o}$ RR-9206, https://hal.inria.fr/hal-01886546

[33] A. AYALA, X. CLAEYS, L. GRIGORI.*Linear-time CUR approximation of BEM matrices*, Inria PARIS, October 2018, n$^{o}$ RR-9208, https://hal.inria.fr/hal-01893036

[34] L. GRIGORI, O. TISSOT.*Scalable Linear Solvers based on Enlarged Krylov subspaces with Dynamic Reduction of Search Directions*, Inria Paris ; Laboratoire Jacques-Louis Lions, UPMC, Paris, July 2018, n$^{o}$ RR-9190, p. 1-30, https://hal.inria.fr/hal-01828521

### Other Publications

[35] M. BONAZZOLI, V. DOLEAN, I. G. GRAHAM, E. A. SPENCE, P.-H. TOURNIER.*A two-level domain-decomposition preconditioner for the time-harmonic Maxwell's equations*, February 2018, https://arxiv.org/abs/1705.08138 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01525438

[36] M. BONAZZOLI, V. DOLEAN, I. G. GRAHAM, E. A. SPENCE, P.-H. TOURNIER.*Two-level preconditioners for the Helmholtz equation*, February 2018, https://arxiv.org/abs/1705.08139 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01525424

[37] X. CLAEYS, P. MARCHAND.*Boundary integral multi-trace formulations and Optimised Schwarz Methods*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01921113

[38] J. COULET, I. FAILLE, V. GIRAULT, N. GUY, F. NATAF.*A fully coupled scheme using Virtual Element Method and Finite Volume for poroelasticity*, October 2018, working paper or preprint, https://hal-ifp.archives-ouvertes.fr/hal-01947455

[39] M. GRAFF, M. J. GROTE, F. NATAF, F. ASSOUS.*How to solve inverse scattering problems without knowing the source term: a three-step strategy*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02002931

[40] F. HECHT, S.-M. KABER.*Fractional decomposition of matrices and parallel computing*, September 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01878765

## References in notes

[41] F. ASSOUS, M. KRAY, F. NATAF, E. TURKEL.*Time-reversed absorbing condition: application to inverse problems*, in "Inverse Problems", 2011, vol. 27, n$^{o}$ 6, 065003, http://stacks.iop.org/0266-5611/27/i=6/a=065003

[42] M. J. GROTE, M. KRAY, F. NATAF, F. ASSOUS.*Time-dependent wave splitting and source separation*, in "Journal of Computational Physics", 2017, vol. 330, p. 981 - 996 [*DOI :* 10.1016/J.JCP.2016.10.021], http://www.sciencedirect.com/science/article/pii/S0021999116305198

# Project-Team ANGE

## Numerical Analysis, Geophysics and Environment

IN COLLABORATION WITH: Laboratoire Jacques-Louis Lions (LJLL)

IN PARTNERSHIP WITH:
**CNRS**

**Centre d'expertise des risques, de l'environnement, des mobilités et de l'aménagement**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Earth, Environmental and Energy Sciences**

# Table of contents

# Project-Team ANGE

*Creation of the Team: 2012 November 01, updated into Project-Team: 2014 January 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
A6.2.6. - Optimization
A6.3. - Computation-data interaction
A6.3.2. - Data assimilation
A6.3.4. - Model reduction
A6.3.5. - Uncertainty Quantification

**Other Research Topics and Application Domains:**

B3. - Environment and planet
B3.3. - Geosciences
B3.3.2. - Water: sea & ocean, lake & river
B3.3.3. - Nearshore
B3.4. - Risks
B3.4.1. - Natural risks
B3.4.3. - Pollution
B4. - Energy
B4.3. - Renewable energy production
B4.3.1. - Biofuels
B4.3.2. - Hydro-energy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Jacques Sainte-Marie [Team leader, Inria, Senior Researcher, HDR]
Vivien Mallet [Inria, Researcher]
Martin Parisot [Inria, Researcher]
Yohan Penel [Inria, Researcher]
Julien Salomon [Inria, Senior Researcher, HDR]

**Faculty Members**

Nina Aguillon [Sorbonne Université, Associate Professor]
Edwige Godlewski [Sorbonne Université, Professor, HDR]
Cindy Guichard [Sorbonne Université, Associate Professor]
Boris Haspot [Univ de Dauphine, Associate Professor, HDR]

Anne Mangeney [IPGP, Professor, HDR]

**External Collaborators**

Emmanuel Audusse [Univ. Paris-Nord]
Bernard Di Martino [Univ. de Corse, HDR]
Fabien Souillé [EDF, from Sep 2018]
Pierrick Quémar [Univ. Paris-Nord]

**Technical Staff**

Guillaume Chérel [Inria, until May 2018]
Cédric Doucet [Inria, from Apr 2018]
Jérémy Ledoux [Univ. Paris Dauphine]
Fabien Souillé [Inria, until Sep 2018]

**PhD Students**

Frederic Allaire [Inria]
Léa Boittin [Inria]
Nelly Boulos Al Makary [Univ Paris-Nord, from Oct 2018]
Virgile Dubos [Sorbonne Université]
Ngoc Le [IRSN]
Antoine Lesieur [Inria]
Liudi Lu [Sorbonne Université, from Oct 2018]
Hugo Martin [Sorbonne Université]
Fabien Wahl [Sorbonne Université]

**Post-Doctoral Fellows**

Bilal Al Taki [Inria]
Janelle Hammond [Inria, from Sep 2018]

**Visiting Scientist**

Marie-Odile Bristeau [Retired]

**Administrative Assistant**

Maryse Desnous [Inria]

# 2. Overall Objectives

## 2.1. Presentation

Among all aspects of geosciences, we mainly focus on gravity driven flows arising in many situations such as

- hazardous flows (flooding, rogue waves, landslides...),
- sustainable energies (hydrodynamics-biology coupling, biofuel production, marine energies...),
- risk management and land-use planning (morphodynamic evolutions, early warning systems...)

There exists a strong demand from scientists and engineers in fluid mechanics for models and numerical tools able to simulate not only the water depth and the velocity field but also the distribution and evolution of external quantities such as pollutants or biological species and the interaction between flows and structures (seashores, erosion processes...). The key point of the researches carried out within ANGE is to answer this demand by the development of efficient, robust and validated models and numerical tools.

## 2.2. Scientific challenges

Due to the variety of applications with a wide range of spatial scales, reduced-size models like the shallow water equations are generally required. From the modelling point of view, the main issue is to describe the behaviour of the flow with a reduced-size model taking into account several physical processes such as non-hydrostatic terms, biological species evolution, topography and structure interactions within the flow. The mathematical analysis of the resulting model do not enter the field of hyperbolic equations anymore and new strategies have to be proposed. Moreover, efficient numerical resolutions of reduced-size models require particular attention due to the different time scales of the processes and in order to recover physical properties such as positivity, conservativity, entropy dissipation and equilibria.

The models can remain subject to uncertainties that originate from incomplete description of the physical processes and from uncertain parameters. Further development of the models may rely on the assimilation of observational data and the uncertainty quantification of the resulting analyses or forecasts.

# 3. Research Program

## 3.1. Overview

The research activities carried out within the ANGE team strongly couple the development of methodological tools with applications to real–life problems and the transfer of numerical codes. The main purpose is to obtain new models adapted to the physical phenomena at stake, identify the main properties that reflect the physical meaning of the models (uniqueness, conservativity, entropy dissipation, ...), propose effective numerical methods to approximate their solution in complex configurations (multi-dimensional, unstructured meshes, well-balanced, ...) and to assess the results with data in the purpose of potentially correcting the models.

The difficulties arising in gravity driven flow studies are threefold.

- Models and equations encountered in fluid mechanics (typically the free surface Navier-Stokes equations) are complex to analyze and solve.
- The underlying phenomena often take place over large domains with very heterogeneous length scales (size of the domain, mean depth, wave length, ...) and distinct time scales, *e.g.* coastal erosion, propagation of a tsunami, ...
- These problems are multi-physics with strong couplings and nonlinearities.

## 3.2. Modelling and analysis

Hazardous flows are complex physical phenomena that can hardly be represented by shallow water type systems of partial differential equations (PDEs). In this domain, the research program is devoted to the derivation and analysis of reduced complexity models compared to the Navier-Stokes equations, but relaxing the shallow water assumptions. The main purpose is then to obtain models well-adapted to the physical phenomena at stake.

Even if the resulting models do not strictly belong to the family of hyperbolic systems, they exhibit hyperbolic features: the analysis and discretisation techniques we intend to develop have connections with those used for hyperbolic conservation laws. It is worth noticing that the need for robust and efficient numerical procedures is reinforced by the smallness of dissipative effects in geophysical models which therefore generate singular solutions and instabilities.

On the one hand, the derivation of the Saint-Venant system from the Navier-Stokes equations is based on two approximations (the so-called shallow water assumptions), namely

- the horisontal fluid velocity is well approximated by its mean value along the vertical direction,
- the pressure is hydrostatic or equivalently the vertical acceleration of the fluid can be neglected compared to the gravitational effects.

As a consequence the objective is to get rid of these two assumptions, one after the other, in order to obtain models accurately approximating the incompressible Euler or Navier-Stokes equations.

On the other hand, many applications require the coupling with non-hydrodynamic equations, as in the case of micro-algae production or erosion processes. These new equations comprise non-hyperbolic features and a special analysis is needed.

### 3.2.1. Multilayer approach

As for the first shallow water assumption, *multi-layer* systems were proposed to describe the flow as a superposition of Saint-Venant type systems [30], [33], [34]. Even if this approach has provided interesting results, layers are considered separate and non-miscible fluids, which implies strong limitations. That is why we proposed a slightly different approach [31], [32] based on a Galerkin type decomposition along the vertical axis of all variables and leading, both for the model and its discretisation, to more accurate results.

A kinetic representation of our multilayer model allows to derive robust numerical schemes endowed with crucial properties such as: consistency, conservativity, positivity, preservation of equilibria, ... It is one of the major achievements of the team but it needs to be analyzed and extended in several directions namely:

- The convergence of the multilayer system towards the hydrostatic Euler system as the number of layers goes to infinity is a critical point. It is not fully satisfactory to have only formal estimates of the convergence and sharp estimates would provide an optimal number of layers.
- The introduction of several source terms due for instance to the Coriolis force or extra terms from changes of coordinates seems necessary. Their inclusion should lead to substantial modifications of the numerical scheme.
- Its hyperbolicity has not yet been proven and conversely the possible loss of hyperbolicity cannot be characterised. Similarly, the hyperbolic feature is essential in the propagation and generation of waves.

### 3.2.2. Non-hydrostatic models

The hydrostatic assumption consists in neglecting the vertical acceleration of the fluid. It is considered valid for a large class of geophysical flows but is restrictive in various situations where the dispersive effects (like wave propagation) cannot be neglected. For instance, when a wave reaches the coast, bathymetry variations give a vertical acceleration to the fluid that strongly modifies the wave characteristics and especially its height.

Processing an asymptotic expansion (w.r.t. the aspect ratio for shallow water flows) into the Navier-Stokes equations, we obtain at the leading order the Saint-Venant system. Going one step further leads to a vertically averaged version of the Euler/Navier-Stokes equations involving some non-hydrostatic terms. This model has several advantages:

- it admits an energy balance law (that is not the case for most dispersive models available in the literature),
- it reduces to the Saint-Venant system when the non-hydrostatic pressure term vanishes,
- it consists in a set of conservation laws with source terms,
- it does not contain high order derivatives.

### 3.2.3. Multi-physics modelling

The coupling of hydrodynamic equations with other equations in order to model interactions between complex systems represents an important part of the team research. More precisely, three multi-physics systems are investigated. More details about the industrial impact of these studies are presented in the following section.

- To estimate the risk for infrastructures in coastal zones or close to a river, the resolution of the shallow water equations with moving bathymetry is necessary. The first step consisted in the study of an additional equation largely used in engineering science: The Exner equation. The analysis enabled to exhibit drawbacks of the coupled model such as the lack of energy conservation or the strong variations of the solution from small perturbations. A new formulation is proposed to avoid

these drawbacks. The new model consists in a coupling between conservation laws and an elliptic equation, like the Euler/Poisson system, suggesting to use well-known strategies for the analysis and the numerical resolution. In addition, the new formulation is derived from classical complex rheology models and allowed physical phenomena like threshold laws.

- Interaction between flows and floating structures is the challenge at the scale of the shallow water equations. This study requires a better understanding of the energy exchanges between the flow and the structure. The mathematical model of floating structures is very hard to solve numerically due to the non-penetration condition at the interface between the flow and the structure. It leads to infinite potential wave speeds that could not be solved with classical free surface numerical schemes. A relaxation model was derived to overcome this difficulty. It represents the interaction with the floating structure with a free surface model-type.
- If the interactions between hydrodynamics and biology phenomena are known through laboratory experiments, it is more difficult to predict the evolution, especially for the biological quantities, in a real and heterogeneous system. The objective is to model and reproduce the hydrodynamics modifications due to forcing term variations (in time and space). We are typically interested in phenomena such as eutrophication, development of harmful bacteria (cyanobacteria) and upwelling phenomena.

### *3.2.4. Data assimilation and inverse modelling*

In environmental applications, the most accurate numerical models remain subject to uncertainties that originate from their parameters and shortcomings in their physical formulations. It is often desirable to quantify the resulting uncertainties in a model forecast. The propagation of the uncertainties may require the generation of ensembles of simulations that ideally sample from the probability density function of the forecast variables. Classical approaches rely on multiple models and on Monte Carlo simulations. The applied perturbations need to be calibrated for the ensemble of simulations to properly sample the uncertainties. Calibrations involve ensemble scores that compare the consistency between the ensemble simulations and the observational data. The computational requirements are so high that designing fast surrogate models or metamodels is often required.

In order to reduce the uncertainties, the fixed or mobile observations of various origins and accuracies can be merged with the simulation results. The uncertainties in the observations and their representativeness also need to be quantified in the process. The assimilation strategy can be formulated in terms of state estimation or parameter estimation (also called inverse modelling). Different algorithms are employed for static and dynamic models, for analyses and forecasts. A challenging question lies in the optimization of the observational network for the assimilation to be the most efficient at a given observational cost.

## 3.3. Numerical analysis

### *3.3.1. Non-hydrostatic scheme*

The main challenge in the study of the non-hydrostatic model is to design a robust and efficient numerical scheme endowed with properties such as: positivity, wet/dry interfaces treatment, consistency. It must be noticed that even if the non-hydrostatic model looks like an extension of the Saint-Venant system, most of the known techniques used in the hydrostatic case are not efficient as we recover strong difficulties encountered in incompressible fluid mechanics due to the extra pressure term. These difficulties are reinforced by the absence of viscous/dissipative terms.

### *3.3.2. Space decomposition and adaptive scheme*

In the quest for a better balance between accuracy and efficiency, a strategy consists in the adaptation of models. Indeed, the systems of partial differential equations we consider result from a hierarchy of simplifying assumptions. However, some of these hypotheses may turn out to be irrelevant locally. The adaptation of models thus consists in determining areas where a simplified model (*e.g.* shallow water type) is valid and where it is not. In the latter case, we may go back to the "parent" model (*e.g.* Euler) in the corresponding area.

This implies to know how to handle the coupling between the aforementioned models from both theoretical and numerical points of view. In particular, the numerical treatment of transmission conditions is a key point. It requires the estimation of characteristic values (Riemann invariant) which have to be determined according to the regime (torrential or fluvial).

### *3.3.3. Asymptotic-Preserving scheme for source terms*

Hydrodynamic models comprise advection and sources terms. The conservation of the balance between source terms, typically viscosity and friction, has a significant impact since the overall flow is generally a perturbation around an equilibrium. The design of numerical schemes able to preserve such balances is a challenge from both theoretical and industrial points of view. The concept of Asymptotic-Preserving (AP) methods is of great interest in order to overcome these issues.

Another difficulty occurs when a term, typically related to the pressure, becomes very large compared to the order of magnitude of the velocity. At this regime, namely the so-called *low Froude* (shallow water) or *low Mach* (Euler) regimes, the difference between the speed of the gravity waves and the physical velocity makes classical numerical schemes inefficient: firstly because of the error of truncation which is inversely proportional to the small parameters, secondly because of the time step governed by the largest speed of the gravity wave. AP methods made a breakthrough in the numerical resolution of asymptotic perturbations of partial-differential equations concerning the first point. The second one can be fixed using partially implicit scheme.

### *3.3.4. Multi-physics models*

Coupling problems also arise within the fluid when it contains pollutants, density variations or biological species. For most situations, the interactions are small enough to use a splitting strategy and the classical numerical scheme for each sub-model, whether it be hydrodynamic or non-hydrodynamic.

The sediment transport raises interesting issues from a numerical aspect. This is an example of coupling between the flow and another phenomenon, namely the deformation of the bottom of the basin that can be carried out either by bed load where the sediment has its own velocity or suspended load in which the particles are mostly driven by the flow. This phenomenon involves different time scales and nonlinear retroactions; hence the need for accurate mechanical models and very robust numerical methods. In collaboration with industrial partners (EDF–LNHE), the team already works on the improvement of numerical methods for existing (mostly empirical) models but our aim is also to propose new (quite) simple models that contain important features and satisfy some basic mechanical requirements. The extension of our 3D models to the transport of weighted particles can also be here of great interest.

### *3.3.5. Optimisation*

Numerical simulations are a very useful tool for the design of new processes, for instance in renewable energy or water decontamination. The optimisation of the process according to a well-defined objective such as the production of energy or the evaluation of a pollutant concentration is the logical upcoming challenge in order to propose competitive solutions in industrial context. First of all, the set of parameters that have a significant impact on the result and on which we can act in practice is identified. Then the optimal parameters can be obtained using the numerical codes produced by the team to estimate the performance for a given set of parameters with an additional loop such as gradient descent or Monte Carlo method. The optimisation is used in practice to determine the best profile for turbine pales, the best location for water turbine implantation, in particular for a farm.

# 4. Application Domains

## 4.1. Overview

Sustainable development and environment preservation have a growing importance and scientists have to address difficult issues such as: management of water resources, renewable energy production, bio/geo-chemistry of oceans, resilience of society w.r.t. hazardous flows, urban pollutions, ...

As mentioned above, the main issue is to propose models of reduced complexity, suitable for scientific computing and endowed with stability properties (continuous and/or discrete). In addition, models and their numerical approximations have to be confronted with experimental data, as analytical solutions are hardly accessible for these problems/models. A. Mangeney (IPGP) and N. Goutal (EDF) may provide useful data.

## 4.2. Geophysical flows

Reduced models like the shallow water equations are particularly well-adapted to the modelling of geophysical flows since there are characterized by large time or/and space scales. For long time simulations, the preservation of equilibria is essential as global solutions are a perturbation around them. The analysis and the numerical preservation of non-trivial equilibria, more precisely when the velocity does not vanish, are still a challenge. In the fields of oceanography and meteorology, the numerical preservation of the so-called geostrophic state, which is the balance between the gravity field and the Coriolis force, can significantly improve the forecasts. In addition, data assimilation is required to improve the simulations and correct the dissipative effect of the numerical scheme.

The sediment transport modelling is of major interest in terms of applications, in particular to estimate the sustainability of facilities with silt or scour, such as canals and bridges. Dredging or filling-up operations are expensive and generally not efficient in the long term. The objective is to determine a configuration almost stable for the facilities. In addition, it is also important to determine the impact of major events like emptying dam which is aimed at evacuating the sediments in the dam reservoir and requires a large discharge. However, the downstream impact should be measured in terms of turbidity, river morphology and flood.

## 4.3. Hydrological disasters

It is a violent, sudden and destructive flow. Between 1996 and 2005, nearly 80% of natural disasters in the world have meteorological or hydrological origines. The main interest of their study is to predict the areas in which they may occur most probably and to prevent damages by means of suitable amenities. In France, floods are the most recurring natural disasters and produce the worst damages. For example, it can be a cause or a consequence of a dam break. The large surface they cover and the long period they can last require the use of reduced models like the shallow water equations. In urban areas, the flow can be largely impacted by the debris, in particular cars, and this requires fluid/structure interactions be well understood. Moreover, underground flows, in particular in sewers, can accelerate and amplify the flow. To take them into account, the model and the numerical resolution should be able to treat the transition between free surface and underground flows.

Tsunamis are another hydrological disaster largely studied. Even if the propagation of the wave is globally well described by the shallow water model in oceans, it is no longer the case close to the epicenter and in the coastal zone where the bathymetry leads to vertical accretions and produces substantial dispersive effects. The non-hydrostatic terms have to be considered and an efficient numerical resolution should be induced.

While viscous effects can often be neglected in water flows, they have to be taken into account in situations such as avalanches, debris flows, pyroclastic flows, erosion processes, *...i.e.* when the fluid rheology becomes more complex. Gravity driven granular flows consist of solid particles commonly mixed with an interstitial lighter fluid (liquid or gas) that may interact with the grains and decrease the intensity of their contacts, thus reducing energy dissipation and favoring propagation. Examples include subaerial or subaqueous rock avalanches (e.g. landslides).

## 4.4. Biodiversity and culture

Nowadays, simulations of the hydrodynamic regime of a river, a lake or an estuary, are not restricted to the determination of the water depth and the fluid velocity. They have to predict the distribution and evolution of external quantities such as pollutants, biological species or sediment concentration.

The potential of micro-algae as a source of biofuel and as a technological solution for CO2 fixation is the subject of intense academic and industrial research. Large-scale production of micro-algae has potential for biofuel applications owing to the high productivity that can be attained in high-rate raceway ponds. One of the key challenges in the production of micro-algae is to maximize algae growth with respect to the exogenous energy that must be used (paddlewheel, pumps, ...). There is a large number of parameters that need to be optimized (characteristics of the biological species, raceway shape, stirring provided by the paddlewheel). Consequently our strategy is to develop efficient models and numerical tools to reproduce the flow induced by the paddlewheel and the evolution of the biological species within this flow. Here, mathematical models can greatly help us reduce experimental costs. Owing to the high heterogeneity of raceways due to gradients of temperature, light intensity and nutrient availability through water height, we cannot use depth-averaged models. We adopt instead more accurate multilayer models that have recently been proposed. However, it is clear that many complex physical phenomena have to be added to our model, such as the effect of sunlight on water temperature and density, evaporation and external forcing.

Many problems previously mentioned also arise in larger scale systems like lakes. Hydrodynamics of lakes is mainly governed by geophysical forcing terms: wind, temperature variations, ...

## 4.5. Sustainable energy

One of the booming lines of business is the field of renewable and decarbonated energies. In particular in the marine realm, several processes have been proposed in order to produce electricity thanks to the recovering of wave, tidal and current energies. We may mention water-turbines, buoys turning variations of the water height into electricity or turbines motioned by currents. Although these processes produce an amount of energy which is less substantial than in thermal or nuclear power plants, they have smaller dimensions and can be set up more easily.

The fluid energy has kinetic and potential parts. The buoys use the potential energy whereas the water-turbines are activated by currents. To become economically relevant, these systems need to be optimized in order to improve their productivity. While for the construction of a harbour, the goal is to minimize swell, in our framework we intend to maximize the wave energy.

This is a complex and original issue which requires a fine model of energy exchanges and efficient numerical tools. In a second step, the optimisation of parameters that can be changed in real-life, such as bottom bathymetry and buoy shape, must be studied. Eventually, physical experiments will be necessary for the validation.

## 4.6. Urban environment

The urban environment is essentially studied for air and noise pollutions. Air pollution levels and noise pollution levels vary a lot from one street to next. The simulations are therefore carried out at street resolution and take into account the city geometry. The associated numerical models are subject to large uncertainties. Their input parameters, e.g. pollution emissions from road traffic, are also uncertain. Quantifying the simulation uncertainties is challenging because of the high computational costs of the numerical models. An appealing approach in this context is the use of metamodels, from which ensembles of simulations can be generated for uncertainty quantification.

The simulation uncertainties can be reduced by the assimilation of fixed and mobile sensors. High-quality fixed monitoring sensors are deployed in cities, and an increasing number of mobile sensors are added to the observational networks. Even smartphones can be used as noise sensors and dramatically increase the spatial coverage of the observations. The processing and assimilation of the observations raises many questions regarding the quality of the measurements and the design of the network of sensors.

## 4.7. SmartCity

There is a growing interest for environmental problems at city scale, where a large part of the population is concentrated and where major pollutions can occur. Numerical simulation is well established to study the urban environment, *e.g.* for road traffic modelling. As part of the smartcity movement, an increasing number of sensors collect measurements, at traditional fixed observation stations, but also on mobile devices, like smartphones. They must properly be taken into account given their number but also their potential low quality.

Pratical applications include air pollution and noise pollution. These directly relate to road traffic. Data assimilation and uncertainty propagation are key topics in these applications.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

**Human resources**

A major event in the year was new positions of J. Sainte-Marie (Détachement at Inria, 2 years position) and of Y. Penel (Advanced Research Position, 3 years position). Two new students have started a PhD (Liudi Lu and Nelly Boulos Al Makary).

**Evaluation of the team**

This year, the team went through the first evaluation since its creation. The report was very positive, as this excerpt shows:

*The activity of the team in modeling and mathematical and numerical analysis has lead to significant contributions in various areas. In particular, we mention the study of models that can reproduce specific 'dispersive effects,' observed in nature, or the review of several multi-physics models that incorporate the coupling of heterogeneous systems. The theoretical analysis of the models has often led to the proposal of new algorithmic developments and new numerical techniques and, in general, it has resulted in a significant advancement of scientific knowledge.*

**Scientific activities** There has been major achievements within the team in the framework of dispersive models.

As detailed in Section 10.1, members of the team were involved in the organisation of a substantial number of scientific events, either in the framework of national initiatives or due to the expertise in the field. Members are is particularly involved in the mathematical community.

### *5.1.1. Awards*

- Léa Boittin received the award of the best presentation at GDR-EGRIN summer school in June,
- Léa Boittin was rewarded by Best Phd Student Poster Award, at CMWR XXII, Saint-Malo,
- Janelle Hammond received a post-doctoral grant from DIM Math Innov 2018.

# 6. New Software and Platforms

## 6.1. Freshkiss

*FREe Surface Hydrodynamics using KInetic SchemeS*
KEYWORDS: Finite volume methods - Hydrostatic Navier-Stokes equations - Free surface flows

FUNCTIONAL DESCRIPTION: Freshkiss3D is a numerical code solving the 3D hydrostatic and incompressible Navier-Stokes equations with variable density.

- Participants: Fabien Souille, Emmanuel Audusse, Jacques Sainte Marie and Marie-Odile Bristeau
- Partners: UPMC - CEREMA
- Contact: Jacques Sainte Marie

## 6.2. TSUNAMATHS

KEYWORDS: Modeling - Tsunamis
FUNCTIONAL DESCRIPTION: Tsunamaths is an educational platform aiming at simulating historical tsunamis. Real data and mathematical explanations are provided to enable people to better understand the overall process of tsunamis.

- Participants: Emmanuel Audusse, Jacques Sainte Marie and Raouf Hamouda
- Contact: Jacques Sainte Marie
- URL: http://tsunamath.paris.inria.fr/

## 6.3. Verdandi

KEYWORDS: HPC - Model - Software Components - Partial differential equation
FUNCTIONAL DESCRIPTION: Verdandi is a free and open-source (LGPL) library for data assimilation. It includes various such methods for coupling one or several numerical models and observational data. Mainly targeted at large systems arising from the discretization of partial differential equations, the library is devised as generic, which allows for applications in a wide range of problems (biology and medicine, environment, image processing, etc.). Verdandi also includes tools to ease the application of data assimilation, in particular in the management of observations or for a priori uncertainty quantification. Implemented in C++, the library may be used with models implemented in Fortran, C, C++ or Python.

- Participants: Dominique Chapelle, Gautier Bureau, Nicolas Claude, Philippe Moireau and Vivien Mallet
- Contact: Vivien Mallet
- URL: http://verdandi.gforge.inria.fr/

## 6.4. Polyphemus

KEYWORD: Simulation
FUNCTIONAL DESCRIPTION: Polyphemus is a modeling system for air quality. As such, it is designed to yield up-to-date simulations in a reliable framework: data assimilation, ensemble forecast and daily forecasts. Its completeness makes it suitable for use in many applications: photochemistry, aerosols, radionuclides, etc. It is able to handle simulations from local to continental scales, with several physical models. It is divided into three main parts:

libraries that gather data processing tools (SeldonData), physical parameterizations (AtmoData) and post-processing abilities (AtmoPy),

programs for physical pre-processing and chemistry-transport models (Polair3D, Castor, two Gaussian models, a Lagrangian model),

model drivers and observation modules for model coupling, ensemble forecasting and data assimilation.

- Participants: Sylvain Doré and Vivien Mallet
- Contact: Vivien Mallet
- URL: http://cerea.enpc.fr/polyphemus/

## 6.5. Urban noise analysis

KEYWORD: Environment perception
FUNCTIONAL DESCRIPTION: This software processes mobile observations collected by the application Ambiciti (previously known as SoundCity). It can merge simulated noise maps with the mobile observations.

- Authors: Raphaël Ventura, Vivien Mallet and Guillaume Cherel
- Contact: Vivien Mallet

## 6.6. Freshkiss3D

- Authors: Jacques Sainte Marie, Marie-Odile Bristeau, Anne-Céline Boulanger, Raouf Hamouda, Emmanuel Audusse, Alain Dervieux, Bijan Mohammadi and David Froger
- Partner: UPMC
- Contact: Jacques Sainte Marie

# 7. New Results

## 7.1. Numerical methods for fluid flows

**Participants:** Jacques Sainte-Marie, Virgile Dubos, Cindy Guichard, Martin Parisot, Marie-Odile Bristeau, Fabien Souillé, Edwige Godlewski, Yohan Penel.

### *7.1.1. Advancing dynamical cores of oceanic models across all scales*

Oceanic numerical models are used to understand and predict a wide range of processes from global paleoclimate scales to short-term prediction in estuaries and shallow coastal areas. One of the overarching challenges, and the main topic of the COMMODORE workshop, is the appropriate design of the dynamical cores given the wide variety of scales of interest and their interactions with atmosphere, sea-ice, biogeochemistry, and even societal processes. The construction of a dynamical core is a very long effort which takes years and decades of research and development and which requires a collaborative mixture of scientific disciplines. In [14], we present a significant number of fundamental choices, such as which equations to solve, which horizontal and vertical grid arrangement is adequate, which discrete algorithms allows jointly computational efficiency and sufficient accuracy, etc.

### *7.1.2. A Well-balanced Finite Volume Scheme for Shallow Water Equations with Porosity*

Our work [20] aims to study the ability of a single porosity-based shallow water model to modelize the impact of vegetation in open-channel flows. More attention on flux and source terms discretizations are required in order to archive the well-balancing and shock capturing. We present a new Godunov-type finite volume scheme based on a simple-wave approximation and compare it with some other methods in the literature. A first application with experimental data was performed.

### *7.1.3. The gradient discretisation method*

The monograph [21] is dedicated to the presentation of the gradient discretisation method (GDM) and to some of its applications. It is intended for masters students, researchers and experts in the field of the numerical analysis of partial differential equations.The GDM is a framework which contains classical and recent discretisation schemes for diffusion problems of different kinds: linear or non-linear, steady-state or time-dependent.

### 7.1.4. Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow

This work [29] is devoted to the numerical resolution in multidimensional framework of a hierarchy of reduced models of the free surface Euler equations. In a first part, entropy-satisfying scheme is proposed fo the monolayer dispersive model [Green, Naghdi '76] and [Bristeau, Mangeney, Sainte-Marie, Seguin '15]. In a second part, the strategy is extended to the layerwise models proposed in [Fernandez-Nieto, Parisot, Penel, Sainte-Marie]. To illustrate the accuracy and the robustness of the strategy, several numerical experiments are performed. In particular, the strategy is able to deal with dry areas without particular treatment.

### 7.1.5. Numerical approximation of the 3d hydrostatic Navier-Stokes system with free surface

In this work [23], we propose a stable and robust strategy to approximate the 3d incompressible hydrostatic Euler and Navier-Stokes systems with free surface. Compared to shallow water approximation of the Navier-Stokes system, the idea is to use a Galerkin type approximation of the velocity field with piecewise constant basis functions in order to obtain an accurate description of the vertical profile of the horizontal velocity.

### 7.1.6. Congested shallow water model: floating object

In [27], we are interested in the floating body problem on a large space scale. We focus on objects floating freely in the water such as icebergs or wave energy converters. The formulation of the fluid-solid interaction using the congested shallow water model for the fluid and Newton's second law of motion for the solid is given and a strong coupling between the two systems is explained. The energy transfer between the solid and the water is focused on since it is of major interest for energy production. A numerical resolution based on the coupling of a finite volume scheme for the fluid and a Newmark scheme for the solid is presented. An entropy correction based on an adapted choice of discretization for the coupling terms is made in order to ensure a dissipation law at the discrete level. Simulations are presented to validate the method and to show the feasibility of more complex cases.

### 7.1.7. Numerical strategies for a dispersive layer-averaged model

A hierarchy of models has been derived in [12] to approximate the Euler equations by means of a layer-averaging procedure. This results in several dispersive models with one velocity field per layer. The structure of the equations induces issues of efficiency. The standard splitting between hydrostatic and non-hydrostatic components leads to a prohibitive computational costs. In a work in progress, we are investigating a new strategy to solve the projection step in a cheaper way. This is assessed by means of steady nontrivial solutions of the dispersive equations.

### 7.1.8. Methods of Reflections

The basic idea of the method of reflections appeared almost two hundred years ago; it is a method of successive approximations for the interaction of particles within a fluid, and it seems intuitively related to the Schwarz domain decomposition methods, the subdomains being the complements of the particle domains. We show in [25] that indeed there is a direct correspondence between the methods of reflections and Schwarz methods in the two particle/subdomain case. This allows us to give a new convergence analysis based on maximum principle techniques with precise convergence estimates that one could not obtain otherwise. We then show however also that in the case of more than two particles/subdomains, the methods of reflections and the Schwarz methods are really different methods, with different convergence properties. We finally also introduce for the first time coarse corrections for the methods of reflections to make them scalable in the case when the number of particles becomes large.

## 7.2. Modelling

**Participants:** Marie-Odile Bristeau, Jacques Sainte-Marie, Fabien Souillé, Emmanuel Audusse, Léa Boitin, Martin Parisot, Di Martino Bernard, Anne Mangeney.

### *7.2.1. How do microalgae perceive light in a high-rate pond? Towards more realistic Lagrangian experiments*

In [10], we present a multidisciplinary downscaling study, where we first reconstructed single cell trajectories in an open raceway using an original hydrodynamical model offering a powerful discretization of the Navier–Stokes equations tailored to systems with free surfaces. The trajectory of a particular cell was selected and the associated high-frequency light pattern was computed. This light pattern was then experimentally reproduced in an Arduino-driven computer controlled cultivation system with a low density Dunaliella salina culture. The effect on growth and pigment content was recorded for various frequencies of the light pattern, by setting different paddle wheel velocities.

### *7.2.2. Modeling and simulation of sediment transport*

A previous derivation of the sediment layer model has then been extended. Depending on the scaling chosen for the physical parameters, different models are obtained. The model we are interested in is the non-local model (with a viscosity term). Several numerical schemes are implemented and studied to simulate this model. Only one of these schemes is satisfactory. Simulations of the coupled water-sediment systems are made. The influence of the viscosity is emphasized. Turning on the non-local term allows to simulate dune growth and propagation.

Following the previous work, a numerical scheme for the sediment layer is proposed. The numerical scheme is tested. The influence of the viscosity on the behaviour of the sediment layer is studied. A numerical strategy for the resolution of the coupled model (water layer and sediment layer) is implemented. The behaviour of the coupled system is numerically assessed. Academic test cases are performed.

### *7.2.3. The Navier-Stokes system with temperature and salinity for free-surface flows*

We model free surface flows where density variations coming e.g. from temperature or salinity differences play a significant role. Starting from the compressible Navier-Stokes system, a model is derived by performing the incompressible limit (the dependence of the density on the pressure is removed). A layer-averaged formulation of the model is proposed. The layer-averaged model satisfies a dissipative energy balance. A numerical scheme is proposed. It verifies several stability properties (positivity, well-balancing, maximum principle on the density). Numerical simulations are performed. The differences with models relying on the classical Boussinesq approximation are shown.

### *7.2.4. Various analytical solutions for the incompressible Euler and Navier-Stokes systems with free surface*

In this paper [24], we propose several time dependent analytical solutions for the incompressible Euler and Navier-Stokes systems with free surface. The given analytical solutions concerns the hydrostatic and nonhydrostatic Euler and Navier-Stokes systems.

## 7.3. Functional analysis of PDE models in Fluid Mechanics

**Participants:** Bilal Al Taki, Boris Haspot.

### *7.3.1. New functional inequality and its application*

In [22], we prove by simple arguments a new kind of Logarithmic Sobolev inequalities generalizing two known inequalities founded in some papers related to fluid dynamics models. As a by product, we show how our inequality can help in obtaining some important a priori estimates for the solution of the Navier-Stokes-Korteweg system.

### 7.3.2. Vortex solutions for the compressible Navier-Stokes equations with general viscosity coefficients in 1D: regularizing effects or not on the density

We consider Navier-Stokes equations for compressible viscous fluids in the one-dimensional case with general viscosity coefficients. We prove the existence of global weak solution when the initial momentum $\rho_0 u_0$ belongs to the set of the finite measure $\mathcal{M}(\mathbf{R})$ and when the initial density $\rho_0$ is in the set of bounded variation functions $BV(\mathbf{R})$. In particular it allows to deal with initial momentum which are Dirac masses and initial density which admit shocks. We can observe in particular that this type of initial data have infinite energy. Furthermore we show that if the coupling between the density and the velocity is sufficiently strong then the initial density which admits initially shocks is instantaneously regularized and becomes continuous. This coupling is expressed via the regularity of the so called effective velocity $v = u + \frac{\mu(\rho)}{\rho^2}\psi_x\rho$ with $\mu(\rho)$ the viscosity coefficient. Inversely if the coupling between the initial density and the initial velocity is too weak (typically $\rho_0 v_0 \in \mathcal{M}(\mathbf{R})$) then we prove the existence of weak energy in finite time but the density remains a priori discontinuous on the time interval of existence.

### 7.3.3. Strong solution for Korteweg system

In this paper we investigate the question of the local existence of strong solution for the Korteweg system in critical spaces when $N \geq 1$ provided that the initial data are small. More precisely the initial momentum $\rho_0 u_0$ belongs to $\mathrm{bmo}_T^{-1}(\mathbf{R}^N)$ for $T > 0$ and the initial density $\rho_0$ is in $L^\infty(\mathbf{R}^N)$ and far away from the vacuum. This result extends the so called Koch-Tataru theorem for the incompressible Navier-Stokes equations to the case of the Korteweg system. It is also interesting to observe that any initial shock on the density is instantaneously regularized inasmuch as the density becomes Lipschitz for any $\rho(t,\cdot)$ with $t > 0$. We also prove the existence of global strong solution for small initial data $(\rho_0 - 1, \rho_0 u_0)$ in the homogeneous Besov spaces $(\dot{B}^{\mathbf{N}-1}_{2,\infty}(\mathbf{R}^{\mathbf{N}} \cap \dot{B}^{\mathbf{N}}_{2,\infty}(\mathbf{R}^{\mathbf{N}} \cap L^\infty(\mathbf{R}^{\mathbf{N}})) \times (\dot{B}^{\mathbf{N}-1}_{2,\infty}(\mathbf{R}^{\mathbf{N}}))^{\mathbf{N}}$. This result allows in particular to extend in dimension $N = 2$ the notion of Oseen solutions defined for incompressible Navier-Stokes equations to the case of the Korteweg system when the vorticity of the momentum $\rho_0 u_0$ is a Dirac mass $\alpha\delta_0$ with $\alpha$ sufficiently small. However unlike the Navier Stokes equations

## 7.4. Assessments of models by means of experimental data and assimilation

**Participants:** Vivien Mallet, Ngoc Bao Tran Le, Antoine Lesieur, Frédéric Allaire, Hammond Janelle.

### 7.4.1. Uncertainty quantification of on-road traffic emissions

Road traffic emissions of air pollutants depend on both traffic flow and vehicle emission factors. At metropolitan scale, traffic flow can be obtained by traffic assignment models, and emission factors can be computed from the traffic flow using COPERT IV formulas. Global sensitivity analyses, especially the computation of Sobol' indices, was carried out for the traffic model and the air pollutant emissions. In the process, the traffic model was replaced by a metamodel, or surrogate model, in order to reduce the high computational burden. The results identified the most important input parameters, e.g., the demand associated with small travel distances (for the traffic flow) or the gasoline car share (for the air pollutant emissions). Furthermore, the uncertainties in traffic flow and pollutant emissions was quantified by propagating into the model the uncertainties in the input parameters. Large ensembles of traffic flows were generated and evaluated with traffic flow measurements.

### 7.4.2. Uncertainty quantification in atmospheric dispersion of radionuclides

In collaboration with IRSN (Institute of Radiation Protection and Nuclear Safety), we investigated the uncertainties of the atmospheric-dispersion forecasts that are used during an accidental release of radionuclides such as the Fukushima disaster. These forecasts are subject to considerable uncertainties which originate from inaccurate weather forecasts, poorly known source term and modeling shortcomings. In order to quantify the uncertainties, we designed a metamodel and investigated the calibration of the probability distribution of the input variables like the source term or the meteorological variables.

### *7.4.3. Metamodeling corrected by observational data*

An air quality model at urban scale computes the air pollutant concentrations at street resolution based on various emissions, meteorology, imported pollution and city geometry. Because of the computational cost of such model, we previously designed a metamodel using dimension reduction and statistical emulation. Novel work was dedicated to the correction of this metamodel using observational data. The proposed approach builds a corrected metamodel that is still much faster than the original model, but also performs better when compared to new observations.

### *7.4.4. Sensitivity analysis and metamodeling of an urban noise model*

Urban noise mapping models simulate the propagation of noise, originating from emission sources (e.g., road traffic), in all street of a city, based on its geometry. They are subject to uncertainties due to incomplete and erroneous data. We carried out screening studies in order to evaluate the sensitivity of the computed noise to the uncertain data. Further work dealt with the development of a metamodel, which will open the way to uncertainty quantification. The work was carried out with the model NoiseModelling and applied to the noise mapping of Lorient (France).

### *7.4.5. Monte Carlo simulation and ensemble evaluation for wildland fire propagation*

We worked on Monte Carlo simulations of wildland fires. The objective was to evaluate how the uncertainties lying in all the inputs of a fire propagation model can be propagated through the model. A careful review of the literature allowed us to define varying intervals for all the uncertain inputs. The Monte Carlo simulations were then evaluated with ensemble scores, using the observations of the final contours for a number of real cases. The ensemble scores were inspired by classical scores used in meteorology, but were adapted to the nature of the fire observations.

### *7.4.6. Metamodeling of a complete air quality simulation chain*

With the objective of uncertainty quantification, we worked in [15] on the generation of a metamodel for the simulation of urban air quality, using a complete simulation chain including dynamic traffic assignment, the computation of air pollutant emissions and the dispersion of the pollutant in a city. The traffic model and the dispersion model are computationally costly and operate in high dimension. We employed dimension reduction, and coupled it with Kriging in order to build a metamodel for the complete simulation chain.

## 7.5. Software Devlopments

**Participant:** Cédric Doucet.

**Improvements in the FRESHKISS3D code** Several improvements have been achieved in FreshKiss3D :

1. installation step is simpler due to the usage of a YAML file listing third-party libraries;
2. Mac OS is now supported;
3. continuous integration is performed on Ubuntu 16 and OSX with different compilers (GCC, Clang) and different builds (debug, release);
4. a major bug in the computation of fluxes has been fixed;
5. the number of third-party libraries has been minimized (geomalgo, metis4py);
6. build automation is now based on CMake (instead of Waf);
7. documentation has been updated and it is now published during the continuous integration process by means of Gitlab pages;
8. continuous integration has been optimized (better slaves, parallelization)

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

- A contract (2016-2018) has been made (130.000 euros) with SAUR, IAV (Institut d'Aménagement de la Vilaine) and Agence de l'eau Loire-Bretagne in collaboration with SciWorks Technologies. It deals with the modelling and the simulation of chlorides entry in the Vilaine reservoir.
- A part of the ANR project Hyflo-Eflu relies on a collaboration with the company "HydroTube Energie". It comprises the recruitment of a young engineer (J. Ledoux) and regular meetings with industrial (Bordeaux) and academic partners (Nantes). See below for more details about the scientific content of this project.
- A part of the ANR project ESTIMAIR includes the SME NUMTECH for a commercial deployment of the project results. (Bordeaux) and academic partners (Nantes). See below for more details about the scientific content of this project.
- J. Sainte-Marie, C. Guichard, Y. Penel, J. Salomon are part of an agreement between Institut Carnot SMILES (Sorbonne Univ., Thomas Boiveau) and the corporation GTT about the improvement of a modeling tool for gas flows in the isolation spaces of LNG tanks

## 8.2. Bilateral Grants with Industry

P. Quémar's PhD thesis is funded by EDF (CIFRE). His PhD is entitled "3D numerical simulations of environmental hydrolics: application to Telemac".

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR MFG (2016-2021)*

**Participant:** Julien Salomon.

Project acronym: MFG

Project title: Mean Field Games

Coordinator: Sébastien Boyaval (LHSV/ENPC)

Funding: 299 160 euros.

CMean field game theory (MFG) is a new and active field of mathematics, which analyses the dynamics of a very large number of agents. Introduced about ten years ago, MFG models have been used in different fields: economics, finance, social sciences, engineering,... MFG theory is at the intersection of mean field theory, mathematical game theory, optimal control, stochastic analysis, variation calculation, partial differential equations and scientific calculation. Drawing on an internationally recognized French team on the subject, the project seeks to obtain major contributions in 4 main directions: the "medium field" aspect (i.e., how to obtain macroscopic models from microscopic models); the analysis of new MFG systems; their numerical analysis; the development of new applications. In this period of rapid expansion of MFG models, the project seeks to foster French leadership in the field and attract new researchers from related fields.

### *9.1.2. ANR INFAMIE (2015-2019)*

**Participant:** Boris Haspot.

Program: ANR Défi de tous les savoirs (DS10) 2015

Project acronym: INFAMIE

Project title: INhomogeneous Flows : Asymptotic Models and Interfaces Evolution

Coordinator: Raphaël Danchin (Univ. Paris-Est)

Funding: 232 960 euros.

Our project aims at a better mathematical understanding of several models for the evolution of inhomogeneous flows. Through three main lines of research (see below), we will pursue a twofold final objective. First, we want to develop the current theory of regular solutions for several equations for the evolution of fluids, proposing a new approach and developing tools that are likely to be efficient in various areas of PDEs. Second, for a few selected concrete systems that describe flows in the earth environment or in astrophysics, we wish to use this general approach to extract as much information as possible concerning the qualitative behavior of the solutions.

### *9.1.3. ANR SEDIFLO (2015-2019)*

**Participants:** Emmanuel Audusse, Martin Parisot.

Program: ANR Défi 1 “Gestion sobre des ressources et adaptation au changement climatique” (JCJC)

Project acronym: SEDIFLO

Project title: Modelling and simulation of solid transport in rivers

Coordinator: Sébastien Boyaval (LHSV/ENPC)

Based on recent theoretical and experimental results, this project is aimed at modelling transport of sediments within rivers. It will rely on innovations from the point of view of rheology as well as advanced mathematical tools (asymptotic model reduction, PDE discretisation).

### *9.1.4. ANR Hyflo-Eflu (2016-2019)*

**Participants:** Jérémy Ledoux, Martin Parisot, Jacques Sainte-Marie, Julien Salomon.

ANR project call: Energies marines renouvelables

Project acronym: Hyflo-Eflu

Project title: Hydroliennes flottantes et énergie fluviale

Coordinator: Julien Salomon

The project is a collaboration between the Inria-team ANGE, specialist of free surface flow and optimisation, and the industrial developers of the turbine, HYDROTUBE ENERGIE. The objective of the project HyFlo-EFlu is to deliver a numerical software able to simulate the dynamic of a floating water turbine in real context. For the academic partner, the main challenge is in the simulation of the floating structure at the scale of the river, and the modelling of the vertical and horisontal axis turbine. For the industrial partner, the objective is the validation of the stability of the structure and the performance in term of energy production.

### *9.1.5. ANR CHARMS (2016-2020)*

**Participant:** Cindy Guichard.

ANR project call: Transformations et inter-conversions énergétiques

Project acronym: CHARMS

Project title: Modèles de réservoirs quantitatifs pour les systèmes hydrothermaux complexes

Coordinator: Simon Lopez (BRGM)

Funding: 73k euros for LJLL (in 767k euros for the whole project)

CHARMS ANR project is focused on the mathematical methods and software tools dedicated to the simulation of the physical models issued from geothermal engineering. The final objective is the achievement of a highly parallel code, validated on realistic cases.

### 9.1.6. CNRS Mocha (2017-2018)

**Participant:** Martin Parisot.

CNRS project call: LEFE

Project acronym: MOCHA

Project title: Multi-dimensiOnal Coupling in Hydraulics and data Assimilation

Coordinator: Martin Parisot

Funding: 14k euros

*In collaboration with S. Barthélémy, N. Goutal, S. Ricci, M. Hoang Le.*

Multi-dimensionnal coupling in river hydrodynamics offers a conveninent solution to properly model complex flow while limiting the computational cost and making the most of pre-exsiting models. The project aims to adapt the lateral interface coupling proposed in [35] to the implicit version and test it on real data for the Garonne River.

### 9.1.7. Inria Project Lab "Algae in Silico" (2015-2018)

**Participants:** Marie-Odile Bristeau, Yohan Penel, Jacques Sainte-Marie, Fabien Souillé.

In the aftermath of the ADT In@lgae (2013–2015), we developed a simulation tool for microalgae culture. An Inria Project Lab "Algae in Silico" has started in collaboration with Inria teams BIOCORE and DYLISS. It concerns microalgae culture for biofuel production and the aim is to provide an integrated platform for numerical simulation "from genes to industrial processes".

### 9.1.8. Inria Project Lab "CityLab" (2015-2018)

**Participants:** Vivien Mallet, Raphaël Ventura.

CityLab@Inria studies ICT solutions toward smart cities that promote both social and environmental sustainability.

### 9.1.9. GdR EGRIN (2017–2021)

**Participants:** Emmanuel Audusse, Bernard Di Martino, Nicole Goutal, Cindy Guichard, Anne Mangeney, Martin Parisot, Jacques Sainte-Marie.

EGRIN stands for Gravity-driven flows and natural hazards. J. Sainte-Marie is the head of the scientific committee of this CNRS research group and A. Mangeney is a member of the committee. Other members of the team involved in the project are local correspondents. The scientific goals of this project are the modelling, analysis and simulation of complex fluids by means of reduced-complexity models in the framework of geophysical flows.

### 9.1.10. ANR FireCaster (2017-2020)

**Participants:** Frédéric Allaire, Vivien Mallet.

ANR project call: DS0104

Project acronym: FireCaster

Project title: Plateforme de prévision incendie et de réponse d'urgence

Coordinator: Jean-Baptiste Filippi (Univ. Corse)

Funding: 442k euros

The goal of the FireCaster project is to prototype a fire decision support system at the national scale to estimate upcoming fire risk (H+24 to H+48) and in case of crisis, to predict fire front position and local pollution (H+1 to H+12).

### *9.1.11. ANR CENSE (2017-2020)*

**Participants:** Antoine Lesieur, Vivien Mallet.

ANR project call: DS0601

Project acronym: CENSE

Project title: Caractérisation des environnements sonores urbains : vers une approche globale associant données libres, mesures et modélisations

Coordinator: Judicaël Picaut (IFSTTAR)

Funding: 856k euros

The CENSE project aims at proposing a new methodology for the production of more realistic noise maps, based on an assimilation of simulated and measured data through a dense network of low-cost sensors.

### *9.1.12. ANR RAVEX (2017-2020)*

**Participant:** Anne Mangeney.

ANR project call: DS0106

Project acronym: RAVEX

Project title: Développement d'une approche intégrée pour la réduction des Risques Associés au Volcanisme EXplosif, de la recherche sur l'aléa aux outils de gestion de crise : le cas de la Martinique

Coordinator: Olivier Roche (IRD)

Funding: 619k euros

### *9.1.13. ANR CINE-PARA (2015-2019)*

**Participant:** Julien Salomon.

ANR project call: DS0708

Project acronym: CINE-PARA

Project title: Méthodes de parallélisation pour cinétiques complexes

Coordinator: Yvon Maday (LJLL)

## 9.2. European Initiatives

### *9.2.1. FP7 & H2020 Projects*

#### *9.2.1.1. ERC Consolidator Grant (2013-2018)*

**Participants:** Anne Mangeney, Hugo Martin.

The project SLIDEQUAKES is about detection and understanding of landslides by observing and modelling gravitational flows and generated earthquakes and is funded by the European Research Council (2 million euros). More precisely, it deals with the mathematical, numerical and experimental modelling of gravitational flows and generated seismic waves coupled with field measurements to better understand and predict these natural hazards and their link with volcanic, seismic and climatic activities.

*9.2.1.2. EoCoE (2015-2018)*

**Participant:** Vivien Mallet.

Title: Energy oriented Centre of Excellence for computer applications

Program: H2020

Duration: October 2015 - October 2018

Coordinator: Édouard Audit (CEA)

Partners: CEA (Commissariat à l'Énergie Atomique et aux Énergies Alternatives, France), Forschungszentrum Julich (Germany), Max Planck Gesellschaft (Germany), ENEA (Agenzia Nazionale Per le Nuove Tecnologie, l'energia E Lo Sviluppo Economico Sostenibile, Italy), CERFACS (European Centre for Research and Advanced Training in Scientific Computing, France), Instytut Chemii Bioorganicznej Polskiej Akademii Nauk (Poland), Universita Degli Studi di Trento (Italy), Fraunhofer Gesellschaft (Germany), University of Bath (United Kingdom), CYL (The Cyprus Institute, Cyprus), CNR (National Research Council of Italy), Université Libre de Bruxelles (Belgium), BSC (Centro Nacional de Supercomputacion, Spain)

Inria contact: Michel Kern (Serena team)

Abstract: The aim of the projevt is to establish an Energy Oriented Centre of Excellence for computing applications (EoCoE). EoCoE (pronounce "Echo") will use the prodigious potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable and low carbon energy supply. To achieve this goal, we believe that the present revolution in hardware technology calls for a similar paradigm change in the way application codes are designed. EoCoE will assist the energy transition via targeted support to four renewable energy pillars: Meteo, Materials, Water and Fusion, each with a heavy reliance on numerical modelling. These four pillars will be anchored within a strong transversal multidisciplinary basis providing high-end expertise in applied mathematics and HPC. EoCoE is structured around a central Franco-German hub coordinating a pan-European network, gathering a total of 8 countries and 23 teams. Its partners are strongly engaged in both the HPC and energy fields; a prerequisite for the long-term sustainability of EoCoE and also ensuring that it is deeply integrated in the overall European strategy for HPC. The primary goal of EoCoE is to create a new, long lasting and sustainable community around computational energy science. At the same time, EoCoE is committed to deliver high-impact results within the first three years. It will resolve current bottlenecks in application codes, leading to new modelling capabilities and scientific advances among the four user communities; it will develop cutting-edge mathematical and numerical methods, and tools to foster the usage of Exascale computing. Dedicated services for laboratories and industries will be established to leverage this expertise and to foster an ecosystem around HPC for energy. EoCoE will give birth to new collaborations and working methods and will encourage widely spread best practices.

### *9.2.2. Collaborations with Major European Organisations*

*9.2.2.1. CNRS PICS NHML (2017-2019)*

**Participants:** Martin Parisot, Yohan Penel, Jacques Sainte-Marie.

Program: CNRS PICS (projet international de collaboration scientifique)

Project acronym: NHML

Project title: non-hydrostatic multilayer models

Duration: 01/17-12/19

Coordinator: Yohan Penel (Inria)

Other partners: IMUS (Sevilla, Spain)

Other Participants: Enrique Fernández-Nieto (Sevilla), Tomas Morales de Luna (Cordoba)

Funding: 12k euros

Abstract: This collaboration aims at designing a hierarchy of multilayer models with a non-hydrostatic pressure as a discretisation along the vertical axis of the Euler equations. The hierarchy relies on the degree of approximation of the variables discretised with a Discontinuous Galerkin method for the vertical direction. These innovative models will imply a theoretical study and the development of numerical tools in dimensions 1 and 2 before the modelling of other physical phenomena (viscosity effects, ...).

## 9.3. International Initiatives

### *9.3.1. Informal International Partners*

Four collaborations with foreign colleagues are to be mentioned:

- **Spain** - A collaboration with Spanish researchers has been initiated in 2016 to derive accurate models and effecient algorithms for free surface flows including non-hydrostatic effects. ANGE applied in 2018 to the Inria Associate Team programme in order to strenghten the collaboration.
- **USA** A joint work with R. LeVeque (Univ. Seattle) and M. Berger (New York Univ.) consists in modelling the impact of asteroids on the generation of tsunamis.
- **Germany** A collaboration with researchers from the University of Constance has been initiated in 2018 about domain decomposition and identifaction algorithms (G. Ciaramella, S. Volkwein).
- **Hong-Kong** A collaboration with F. Kwok on time parallelization for assimilation algorithm has been initiated in 2018.

## 9.4. International Research Visitors

- Y. Penel spent twice two weeks (May, October)at the university of Sevilla (Spain) to collaborate with E. Fernández-Nieto.

### *9.4.1. Visits of International Scientists*

- G. Ciaramella visited J. Salomon (28.05-01.06) to work on a reduction method for identification problem.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

- Y. Penel and J. Sainte-Marie organised (with E. Fernández-Nieto) the workshop “non-hydrostatic effects in oceanography ” that took place at Sevilla univ. on 15-16th October and that gathered 35 international researchers [0].
- J. Sainte-Marie took part of the organization of the 6th EGRIN summer school that took place at Le Lioran from 18th to 21st of June and that gathered 40 researchers [0].
- J. Sainte-Marie took part of the organization of the Workshop COMMODORE "Community for the numerical modeling of the global, regional and coastal ocean" [0] (17-19/9/18).
- B. Haspot and Y. Penel organise the monthly ANGE seminar [0].
- J. Salomon co-organises the LJLL-Inria meetings [0] (twice a month).
- J. Salomon co-organised the minisymposium "Domain-decomposition methods for integral equation problems " at the 25th International Domain Decomposition Conference, DD XXV, in St. John’s, Newfoundland, Canada, July 23-27, 2018.
- L. Boittin co-organises the Junior Seminar at Inria–Paris.
- M. Parisot and J. Salomon organise a workshop entitled “Scientific computing and optimisation processes for renewable energies” at Inria [0] on January 2018.

[0] https://team.inria.fr/ange/workshop-non-hydrostatic-effects-in-oceanography/
[0] https://indico.math.cnrs.fr/event/3345/overview
[0] https://commodore2018.sciencesconf.org/
[0] https://team.inria.fr/ange/gdt-slides/
[0] https://project.inria.fr/rencontresljll/fr/
[0] https://emrsim2018.sciencesconf.org/

### *10.1.1. Journal*

The summary of the reviewing activities of the team is given in the next table.

| **Member** | **Journal** |
|---|---|
| Julien Salomon | CRAS, SIAM SISC |
| Cindy Guichard | J. Comp. Phys., J. Sci. Comp., CRAS, J. Comp. Math., Num. Math. |
| Jacques Sainte-Marie | M2AN, Computer and Fluids, Ocean Modelling, J. Comp. Phys., J. Sci. Comp. Adances in Comp. Math. |
| Martin Parisot | J. Hydraulic Research, Computers and Fluids, J. Comp. Phys., J. de Math. Pures et Appliquées |
| Vivien Mallet | ANR, Journal of Machine Learning Research |
| Edwige Godlewski | Sinum |

### *10.1.2. Invited and contributed Talks*

| Conference | Location | Month | Members involved |
|---|---|---|---|
| Julien Salomon | 7-th Parallel-in-time Integration Workshop (PinT 2018). | Roscoff, France | 02-05/05/2018 |
| Julien Salomon | 25-th International Conference on Domain Decomposition Methods (DD25). | St-Johns, Canada | 23-27/07/18 |
| Julien Salomon | Séminaire du laboratoire de mathématiques Blaise Pascal. | Clermont-Ferrand, France | 01/02/18 |
| Julien Salomon | Séminaire du CMAP. | Palaiseau, France | 15/5/18 |
| Julien Salomon | Séminaire du Laboratoire J.-L. Lions | Paris, France | 30/03/18 |
| Julien Salomon | Séminaire du groupe d'analyse | Universität Konstanz | 14/11/18 |
| Fabien Wahl | Simulation et Optimisation pour les Energies Marines Renouvelables | Inria Paris | 11/01/18 |
| Fabien Wahl | GdT ANGE | Inria Paris | 10/10/18 |
| Fabien Wahl | EGRIN | Le Grand Lioran | 18/06/18 |
| Bilal Al Taki | Journée interne du LJLL | Paris diderot (P7) | 05/04/18 |
| Bilal Al Taki | Séminaire du laboratoire | Marseille | 20/11/18 |
| Frédéric Allaire | 8th International Conference on Forest Fire Research (ICFFR) | Coimbra, Portugal | 12-16/11/18 |
| Virgile Dubos, Martin Parisot | Non-hydrostatic effects in oceanography | Séville, Espagne | 15-16/10/18 |
| Virgile Dubos | 18th Spanish-French Sch. J-L. Lions, Num. Sim. in Phy. and Eng. | Las Palmas, Espagne | 25-29/06/18 |
| Cindy Guichard | rencontres Inria-LJLL en calcul scientifique | Paris | 05/11 |
| Cindy Guichard | Mini-symposium, CANUM | Cap d'Agde | 29/05 |
| Cindy Guichard | séminaire d'analyse numérique, CEA/DIF | Bruyères-le-Châtel | 03/05/18 |
| Cindy Guichard | journée interne du LJLL | Paris | 05/04/18 |
| Cindy Guichard | séminaire d'analyse numérique des EDP | univ. Orsay | 22/03/18 |
| Yohan Penel | Mini-symposium, CANUM | Cap d'Agde | 29/05/18 |
| Yohan Penel | GDR Manu (poster) | Roscoff, France | 02/07/18 |
| Yohan Penel | Workshop "non-hydrostatic effects in oceanography" | Séville, Espagne | 16/10/18 |
| Jacques Sainte-Marie | Plenary in " Free Surface Flows: from Hydrostatic to Non-Hydrostatic Models" | Saint-Malo | 6/6/18 |
| Jacques Sainte-Marie | Séminaire du laboratoire N. Oresme - université de Caen | Caen | 12/02/18 |
| Léa Boittin | GdT ANGE | Inria Paris | 21/02/18 |
| Léa Boittin | GTT LJLL | LJLL, Paris | 30/1 |
| Léa Boittin | CMWR XXII (poster) | Saint-Malo | 04/06/18 |
| Léa Boittin | EGRIN | Le Grand Lioran | 18/6/18 |
| Martin Parisot | Seminaire d'équipe LEMON | Montpellier | 6/2/18 |
| Martin Parisot | SIMAI-UMI Congres | Wroclaw, Poland | 16-20/09/18 |
| Martin Parisot | Séminaire du Laboratoire Jean Leray | Nantes | 11/10/18 |
| Martin Parisot | Conference Balance laws in fluid mechanics, geophysics, biology | Orléans | 19-22/10/18 |
| Martin Parisot | Séminaire "Math et Eau" | Montpellier | 25/10/18 |
| Martin Parisot | Colloque LEFE | Clermont-Ferrand | 28-30/03/18 |
| Martin Parisot | Workshop Non Hydro, Séville | Séville | 10/18 |
| Jacques Sainte-Marie | Séminaire du laboratoire de | Beyrouth université | 19/12/2018 |

### *10.1.3. Leadership within the Scientific Community*

Yohan Penel is a member of the Council Administration of SMAI (2015-2018).

### *10.1.4. Scientific Expertise*

- Yohan Penel was rewiewer for the grant CNRS INSU LEFE
- Vivien Mallet was reviewer for ANR

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

| Member | Level | Institution | Duration | Type | Topic |
|---|---|---|---|---|---|
| J. Salomon | M2 | Univ. Paris-Dauphine | 15 | CM | Cours de rentrée : méthodes numériques pour les EDP |
| J. Salomon | M2 | Univ. Paris-Dauphine | 30 | CM | Méthodes numériques pour des modèles incluants des EDP |
| L. Lu | L1 | SU | 27.4 | TD | Analyse et algèbre pour les sciences |
| B. Al Taki | L1 | SU | 36 | TD | Analyse et algèbre pour les sciences |
| F. Allaire | L1 | SU | 38.5 | TD | Calculus |
| V. Dubos | L3 | Polytech Sorbonne | 32 | TP | Mathématiques appliquées |
| V. Dubos | M1 | Polytech Sorbonne | 14+22 | CM+TP | Traitement numérique |
| V. Dubos | L3 | Polytech Sorbonne | 10 | TP | Projet d'initiation |
| C. Guichard | M2 | SU | 22 | CM+TD | Méthodes numériques |
| C. Guichard | M1 | SU | 58 | TP | Fondements des méthodes numériques |
| C. Guichard | L3 | SU | 21 | TP | Python |
| C. Guichard | M1 | SU | 18 | CM | Mise en oeuvre de la méthode des éléments finis |
| B. Di Martino | L1 | Univ. Corse | 18 | CM+TD | Techniques mathématiques et physiques pour les sciences de la vie |
| B. Di Martino | L2 | Univ. Corse | 72 | CM+TD+TP | Analyse et TP Python Sage algèbre et analyse |
| B. Di Martino | L3 | Univ. Corse | 54 | CM+TD+TP | Analyse numérique matricielle |
| B. Di Martino | M2 | Univ. Corse | 24 | CM+TD+TP | Modélisation master Gestion Intégrée du Littoral et Valorisation Halieutique |
| Y. Penel | L2 | SU | 12 | CM | Analyse vectorielle et intégrales multiples |
| Y. Penel | M1 | Univ. Paris Descartes | 37.5 | CM+TD+TP | Modélisation déterministe en sciences du vivant |
| J. Sainte-Marie | M1 | IPGP | 40 | CM | Modélisation des écoulements gravitaires |
| J. Sainte-Marie | M2 | IPGP | 30 | CM et TP | Méthodes numériques, appli. géosciences |
| J. Sainte-Marie | M2 | SU | 20 | CM | Méthodes numériques pour les systèmes hyperboliques Applications aux énergies renouvelables |
| N. Boulos | L1 | Paris 13 | 48 | TD | Mathématiques pour le parcours aménagé |
| L. Boittin | L3 | SU | 24,5 | TD | Méthodes numériques pour les EDO |
| M. Parisot | L3 | Polytech | 46 | CM+TP | Méthodes numériques |

### *10.2.2. Supervision*

| Supervisor ANGE | Type | Name | Institution | Time | Title |
|---|---|---|---|---|---|
| JS | PhD | Sebastian Reyes-Riffo | Paris-Dauphine | 2016-2019 | Méthodes numériques pour les énergies marines renouvelables |
| JS, VM | PhD | Antoine Lesieur | Inria | 2017-2020 | Estimation d'état et modélisation inverse appliquées à la pollution sonore en milieu urbain |
| JS | PhD | Nadia Jbili | Paris-Dauphine | 2016-2019 | Contrôle optimal pour la résonance magnétique nucléaire |
| JS, JSM | PhD | Liudi Lu | Inria | 2018-2021 | Approches Lagrangiennes pour la modélisation et l'optimisation du couplage hydrodynamique-photosynthèse |
| JSM, VM | PhD | Frédéric Allaire | Inria | 2017-2020 | Quantification du risque incendie par méta-modélisation de la propagation de feux de forêt |
| YP, CG, JSM | PhD | Virgile Dubos | SU | 2017-2020 | Numerical methods for the elliptic/parabolic parts of non-hydrostatic fluid models |
| BDM, BH | Post Doc | Bilal AL Taki | Inria | 09/17-12/18 | Understanding and modeling the rheology of complex surface flow |
| NA, EA, MP | stage M2 | Nelly BOULOS | Inria | 04-08/2018 | Analyse et simulation de modèles d'écoulement à surface libre intégrés selon la verticale |
| EA, MP, JSM, MOB | PhD | Léa Boittin | Paris 6 | 2015-2019 | Modelling, analysis and efficient numerical resolution for erosion processes, |
| MP, NA, EA, MP | PhD | Nelly BOULOS | Paris 13 | 2018-2021 | Modélisation et simulation numérique de la dynamique d'un acquifère érodable |
| VM | PhD | Ngoc Bao Tran Le | Inria | 2016-2019 | Uncertainty quantification based on model reduction for atmospheric dispersion |
| VM | Post Doc | Janelle Hammond | Inria | 2017-2019 | Uncertainty quantification, metamodeling and data assimilation applied to urban air quality |
| EA, MP,JSM | PhD | Léa Boittin | inria | 2016-2019 | Modelling, analysis and efficient numerical resolution for erosion processes |

JS : J. Salomon, VM : Vivien Mallet, JSM : Jacques Sainte-Marie, YP : Yohan Penel, CG : Cindy Guichard, BDM : Bernard Di Martino, BH : Boris Haspot, NA: Nina Aguillon, EA : Emmanuel Audusse, MP : Martin Parisot, EG : Edwige Godlewski

### *10.2.3. Juries*

| Member | Date | Type (PhD, HdR) | role | Name | Institution | Title |
|---|---|---|---|---|---|---|
| JS | Novembre | PhD | rapporteur | Quentin Ansel | Univ. Dijon, TUM Munich | Optimal control ofinhomogeneous spin ensembles: Applications in NMR and Quantum optics |
| JS | Décembre | PhD | rapporteur | Pierre Terrier | ENPC, Université Paris-Est | Simulations numériques pour la prédiction de l'évolution microstructurale d'alliages ferritiques.Une étude de la dynamique d'amas. |
| JS | Décembre | PhD | rapporteur | Amina Benaceur | ENPC, Université Paris-Est | Réduction de modèles en thermique etmécanique non-linéaires |
| YP | Septembre | comité de mi-thèse | examinateur | Moustoifa Rafiou | Univ. Toulon | Modélisation et simulation numérique d'un écoulement à faible nombre de Mach. Application à un réacteur à eaux pressurisées. |
| JSM | Novembre | PhD | rapporteur | Nicolas Peton | IFPEN et univ. Paris-Saclay | Étude et simulation d'un modèle stratigraphique advecto-diffusif non-linéaire avec frontières mobiles |
| JSM | Janvier | PhD | président | Charles Demay | EDF et univ. Savoie Mont-Blanc | Modelling and simulation of transient air-water two-phase flows in hydraulic pipes |
| EG | mars | PhD | membre | Alexis Marboeuf | Ecole Polytechnique | Schémas ALE multi-matériaux totalement conservatifs pour l'hudrodynamique |
| EG | juillet | PhD | présidente | Camilla Fiorini | UVSQ | Analyse de sensibilité pour systèmes hyperboliques non linéaires |
| EG | octobre | PhD | présidente | Julie Llobell | UCA (Nice) | Schémas Volumes Finis à mailles décalées pour la dynamique des gaz |
| EG | novembre | PhD | présidente | Nicolas Cagnart | SU | Quelques approches non linéaires en réduction de complexité |
| EG | novembre | PhD | rapporteure | David Iampetro | AMU (EDF Saclay) | Contribution à la simulation |

JS : J. Salomon, VM : Vivien Mallet, JSM : Jacques Sainte-Marie, YP : Yohan Penel, EG : Edwige Godlewski

## 10.3. Popularization

### *10.3.1. Internal or external Inria responsibilities*

Julien Salomon is a member of the "Comité des usagers de la rue Barrault" in view of the move of Paris Inria Center to Rue Barrault.

### *10.3.2. Articles and contents*

Julien Salomon wrote a vulgarization article "Décomposer et itérer pour résoudre un problème " for the CNRS website " Images des mathématiques " (12/2018)

### *10.3.3. Education*

- Edwige Godlewski is the president of the "commission française pour l'enseignement des mathématiques" (CFEM)
- Jacques Sainte-Marie is a member of the "Groupe de travail : Recherche et développement durable" at the French Ministry of Research.

### *10.3.4. Interventions*

- Julien Salomon took part of the "Salon Culture et Jeux mathématiques" - Stand AMIES (Paris, 24/05/18)
- Julien Salomon gave a talk at Waterford Kamhlaba United World College, Mbabane, Eswatini (Swaziland, 5/8/2018).
- Julien Salomon took part of "Fête de la science", and gave a talk at école rue st-Isaure, 18ème (Paris,11/10/2018)
- Léa Boittin 25/5/18 took part of the "Salon Culture et Jeux mathématiques" - Stand AMIES ( Paris, 25/05/18)
- Vivien Mallet gave a talk "Bruit dans la ville", for the association "Versailles environnement initiative" (01/12/18)

### *10.3.5. Internal action*

Julien Salomon gave talk (3.4.18) at the internal meeting "La demi-heure de science": "Décomposer et itérer pour résoudre un problème complexe, quelques exemples en calcul scientifique".

Anne Mangeney gave talk (4.9.18) at the internal meeting "La demi-heure de science": "Les ondes sismiques : une mine d'informations sur les risques naturels".

# 11. Bibliography

## Major publications by the team in recent years

[1] E. AUDUSSE, M.-O. BRISTEAU, M. PELANTI, J. SAINTE-MARIE.*Approximation of the hydrostatic Navier-Stokes system for density stratified flows by a multilayer model. Kinetic interpretation and numerical validation*, in "J. Comput. Phys.", 2011, vol. 230, p. 3453-3478, http://dx.doi.org/10.1016/j.jcp.2011.01.042

[2] E. AUDUSSE, M.-O. BRISTEAU, B. PERTHAME, J. SAINTE-MARIE.*A multilayer Saint-Venant system with mass exchanges for Shallow Water flows. Derivation and numerical validation*, in "ESAIM Math. Model. Numer. Anal.", 2011, vol. 45, p. 169-200, http://dx.doi.org/10.1051/m2an/2010036

[3] M.-O. BRISTEAU, A. MANGENEY, J. SAINTE-MARIE, N. SEGUIN.*An energy-consistent depth-averaged Euler system: derivation and properties*, in "Discrete and Continuous Dynamical Systems - Series B", 2015, vol. 20, n° 4, 28

[4] J. SAINTE-MARIE.*Vertically averaged models for the free surface Euler system. Derivation and kinetic interpretation*, in "Math. Models Methods Appl. Sci. (M3AS)", 2011, vol. 21, n° 3, p. 459-490, http://dx.doi.org/10.1142/S0218202511005118

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[5] R. VENTURA.*Estimation of urban noise pollution with the assimilation of mobile measurements*, Université Pierre & Marie Curie - Paris 6, June 2018, https://hal.inria.fr/tel-01910084

[6] F. WAHL.*Modeling and analysis of interactions between free surface flows and floating structures*, Sorbonne Université, December 2018, https://tel.archives-ouvertes.fr/tel-01955798

### Articles in International Peer-Reviewed Journal

[7] P. AUMOND, A. CAN, V. MALLET, B. DE COENSEL, C. RIBEIRO, D. BOTTELDOOREN, C. LAVANDIER.*Kriging-based spatial interpolation from measurements for sound level mapping in urban areas*, in "Journal of the Acoustical Society of America", 2018, vol. 143, n° 5, p. 2847-2857 [*DOI :* 10.1121/1.5034799], https://hal.archives-ouvertes.fr/hal-01826354

[8] V. BACHELET, A. MANGENEY, J. DE ROSNY, R. TOUSSAINT, M. FARIN.*Elastic wave generated by granular impact on rough and erodible surfaces*, in "Journal of Applied Physics", January 2018, vol. 123, n° 4, 044901 [*DOI :* 10.1063/1.5012979], https://hal.archives-ouvertes.fr/hal-01907610

[9] S. DELLACHERIE, G. FACCANONI, B. GREC, Y. PENEL.*Accurate steam-water equation of state for two-phase flow LMNC model with phase transition*, in "Applied Mathematical Modelling", 2019, vol. 65, p. 207-233 [*DOI :* 10.1016/J.APM.2018.07.028], https://hal.archives-ouvertes.fr/hal-01111730

[10] D. DEMORY, C. COMBE, P. HARTMANN, A. TALEC, E. PRUVOST, R. HAMOUDA, F. SOUILLÉ, P.-O. LAMARE, M.-O. BRISTEAU, J. SAINTE-MARIE, S. RABOUILLE, F. MAIRET, A. SCIANDRA, O. BERNARD.*How do microalgae perceive light in a high-rate pond? Towards more realistic Lagrangian experiments*, in "Royal Society Open Science", May 2018, vol. 5, n° 5, 180523 [*DOI :* 10.1098/RSOS.180523], https://hal.sorbonne-universite.fr/hal-01830067

[11] M. FARIN, A. MANGENEY, J. DE ROSNY, R. TOUSSAINT, P.-T. TRINH.*Link Between the Dynamics of Granular Flows and the Generated Seismic Signal: Insights From Laboratory Experiments*, in "Journal of Geophysical Research : Earth Surface", June 2018, vol. 123, n° 6, p. 1407-1429 [*DOI :* 10.1029/2017JF004296], https://hal.archives-ouvertes.fr/hal-01907646

[12] E. D. FERNANDEZ-NIETO, M. PARISOT, Y. PENEL, J. SAINTE-MARIE.*A hierarchy of dispersive layer-averaged approximations of Euler equations for free surface flows*, in "Communications in Mathematical Sciences", December 2018, vol. 16, n° 5, p. 1169-1202 [*DOI :* 10.4310/CMS.2018.V16.N5.A1], https://hal.archives-ouvertes.fr/hal-01324012

[13] E. GODLEWSKI, M. PARISOT, J. SAINTE-MARIE, F. WAHL.*Congested shallow water model: roof modelling in free surface flow*, in "ESAIM: Mathematical Modelling and Numerical Analysis", November 2018, vol. 52, n$^{o}$ 5, p. 1679 - 1707, https://hal.inria.fr/hal-01368075

[14] F. LEMARIÉ, H. BURCHARD, L. DEBREU, K. KLINGBEIL, J. SAINTE-MARIE.*Advancing dynamical cores of oceanic models across all scales*, in "Bulletin of the American Meteorological Society", November 2018 [*DOI :* 10.1175/BAMS-D-18-0303.1], https://hal.inria.fr/hal-01939057

[15] V. MALLET, A. TILLOY, D. POULET, S. GIRARD, F. BROCHETON.*Meta-modeling of ADMS-Urban by dimension reduction and emulation*, in "Atmospheric Environment", July 2018, vol. 184, p. 37 - 46 [*DOI :* 10.1016/J.ATMOSENV.2018.04.009], https://hal.inria.fr/hal-01909921

[16] A. SERGEANT, V. YASTREBOV, A. MANGENEY, O. CASTELNAU, J.-P. MONTAGNER, E. STUTZMANN.*Numerical modeling of iceberg capsizing responsible for glacial earthquakes*, in "Journal of Geophysical Research : Earth Surface", 2018, vol. 123, p. 3013-3033 [*DOI :* 10.1029/2018JF004768], https://hal.archives-ouvertes.fr/hal-01943050

[17] J. THOREY, C. CHAUSSIN, V. MALLET.*Ensemble forecast of photovoltaic power with online CRPS learning*, in "International Journal of Forecasting", May 2018, vol. 34, n$^{o}$ 4 [*DOI :* 10.1016/J.IJFORECAST.2018.05.007], https://hal.inria.fr/hal-01909827

[18] R. VENTURA, V. MALLET, V. ISSARNY.*Assimilation of mobile phone measurements for noise mapping of a neighborhood*, in "Journal of the Acoustical Society of America", September 2018, vol. 144, n$^{o}$ 3, p. 1279 - 1292 [*DOI :* 10.1121/1.5052173], https://hal.inria.fr/hal-01909933

### Articles in Non Peer-Reviewed Journal

[19] V. MALLET, R. VENTURA, V. ISSARNY, P. G. RAVERDY.*Assimilation d'observations participatives issues de l'application mobile Ambiciti*, in "Acoustique et Techniques : trimestriel d'information des professionnels de l'acoustique", January 2018, https://hal.inria.fr/hal-01909880

### Conferences without Proceedings

[20] M. H. LE, V. DUBOS, M. OUKACINE, N. GOUTAL.*A Well-balanced Finite Volume Scheme for Shallow Water Equations with Porosity Application to Modelling of Open-Channel Flow through Rigid and Emergent Vegetation*, in "River Flow 2018 - 9th International Conference on Fluvial Hydraulics", Lyon-Villeurbanne, France, September 2018, p. 1-7, https://hal.archives-ouvertes.fr/hal-01706777

### Scientific Books (or Scientific Book chapters)

[21] J. DRONIOU, R. EYMARD, T. GALLOUËT, C. GUICHARD, R. HERBIN.*The gradient discretisation method*, Mathématiques et Applications, Springer International Publishing AG, August 2018, vol. 82, Cf file "changelog_hal.pdf" [*DOI :* 10.1007/978-3-319-79042-8], https://hal.archives-ouvertes.fr/hal-01382358

### Other Publications

[22] B. AL TAKI.*New functional inequality and its application*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01905324

[23] S. ALLGEYER, M.-O. BRISTEAU, D. FROGER, R. HAMOUDA, A. MANGENEY, J. SAINTE-MARIE, F. SOUILLÉ, M. VALLÉE.*Numerical approximation of the 3d hydrostatic Navier-Stokes system with free*

*surface*, September 2018, https://arxiv.org/abs/1709.06267 - working paper or preprint, https://hal.inria.fr/hal-01393147

[24] M.-O. BRISTEAU, B. DI MARTINO, A. MANGENEY, J. SAINTE-MARIE, F. SOUILLÉ.*Various analytical solutions for the incompressible Euler and Navier-Stokes systems with free surface*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01831622

[25] G. CIARAMELLA, M. J. GANDER, L. HALPERN, J. SALOMON.*Methods of Reflections: relations with Schwarz methods and classical stationary iterations, scalability and preconditioning*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01930232

[26] R. EYMARD, C. GUICHARD, X. LHÉBRARD.*Convergence of numerical schemes for a conservation equation with convection and degenerate diffusion*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01939144

[27] E. GODLEWSKI, M. PARISOT, J. SAINTE-MARIE, F. WAHL.*Congested shallow water model: floating object*, September 2018, working paper or preprint, https://hal.inria.fr/hal-01871708

[28] B. HASPOT.*Global bmo -1 (R N ) radially symmetric solution for compressible Navier-Stokes equations with initial density in L $\infty$ (R N )*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01976953

[29] M. PARISOT.*Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow*, February 2018, working paper or preprint, https://hal.inria.fr/hal-01242128

## References in notes

[30] E. AUDUSSE.*A multilayer Saint-Venant model : Derivation and numerical validation*, in "Discrete Contin. Dyn. Syst. Ser. B", 2005, vol. 5, n$^{o}$ 2, p. 189-214

[31] E. AUDUSSE, M.-O. BRISTEAU, M. PELANTI, J. SAINTE-MARIE.*Approximation of the hydrostatic Navier-Stokes system for density stratified flows by a multilayer model. Kinetic interpretation and numerical validation*, in "J. Comput. Phys.", 2011, vol. 230, p. 3453-3478, http://dx.doi.org/10.1016/j.jcp.2011.01.042

[32] E. AUDUSSE, M.-O. BRISTEAU, B. PERTHAME, J. SAINTE-MARIE.*A multilayer Saint-Venant system with mass exchanges for Shallow Water flows. Derivation and numerical validation*, in "ESAIM Math. Model. Numer. Anal.", 2011, vol. 45, p. 169-200, http://dx.doi.org/10.1051/m2an/2010036

[33] F. BOUCHUT, V. ZEITLIN.*A robust well-balanced scheme for multi-layer shallow water equations*, in "Discrete Contin. Dyn. Syst. Ser. B", 2010, vol. 13, p. 739-758

[34] M. CASTRO, J. GARCÍA-RODRÍGUEZ, J. GONZÁLEZ-VIDA, J. MACÍAS, C. PARÉS, M. VÁZQUEZ-CENDÓN.*Numerical simulation of two-layer shallow water flows through channels with irregular geometry*, in "J. Comput. Phys.", 2004, vol. 195, n$^{o}$ 1, p. 202–235

[35] N. GOUTAL, M. PARISOT, F. ZAOUI.*A 2D reconstruction for the transverse coupling of shallow water models*, in "Int. J. Numer. Methods Fluids", 2014, vol. 75, n$^{o}$ 11, p. 775–799

# Project-Team ANTIQUE

## Static Analysis by Abstract Interpretation

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Proofs and Verification**

# Table of contents

# Project-Team ANTIQUE

*Creation of the Team: 2014 January 01, updated into Project-Team: 2015 April 01*

**Keywords:**

**Computer Science and Digital Science:**

A2. - Software
A2.1. - Programming Languages
A2.1.1. - Semantics of programming languages
A2.1.7. - Distributed programming
A2.1.12. - Dynamic languages
A2.2.1. - Static analysis
A2.3. - Embedded and cyber-physical systems
A2.3.1. - Embedded systems
A2.3.2. - Cyber-physical systems
A2.3.3. - Real-time systems
A2.4. - Formal method for verification, reliability, certification
A2.4.1. - Analysis
A2.4.2. - Model-checking
A2.4.3. - Proofs
A2.6.1. - Operating systems
A4.4. - Security of equipment and software
A4.5. - Formal methods for security

**Other Research Topics and Application Domains:**

B1.1. - Biology
B1.1.8. - Mathematical biology
B1.1.10. - Systems and synthetic biology
B5.2. - Design and manufacturing
B5.2.1. - Road vehicles
B5.2.2. - Railway
B5.2.3. - Aviation
B5.2.4. - Aerospace
B6.1. - Software industry
B6.1.1. - Software engineering
B6.1.2. - Software evolution, maintenance
B6.6. - Embedded systems

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Xavier Rival [Team leader, Inria, Senior Researcher, HDR]
Vincent Danos [CNRS, Senior Researcher, HDR]
Cezara Drăgoi [Inria, Researcher]

Jérôme Feret [Inria, Researcher]

**Technical Staff**

Ferdinanda Camporesi [Inria, until Apr 2018]
Yves Stan Le Cornec [Inria]

**PhD Students**

Andreea Beica [Ecole Normale Supérieure Paris]
Gaëlle Candel [Keymetrics]
Marc Chevalier [Ecole Normale Supérieure Lyon]
Hugo Illous [Ecole Normale Supérieure Paris]
Huisong Li [Inria, until Feb 2018]
Thibault Suzanne [Ecole Normale Supérieure Paris]
Jiangchao Liu [Inria, until Mar 2018]

**Visiting Scientist**

Pierre Boutillier [Harvard Medical School]

**Administrative Assistant**

Nathalie Gaudechoux [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

Our group focuses on developing *automated* techniques to compute *semantic properties* of programs and other systems with a computational semantics in general. Such properties include (but are not limited to) important classes of correctness properties.

Verifying safety critical systems (such as avionics systems) is an important motivation to compute such properties. Indeed, a fault in an avionics system, such as a runtime error in the fly-by-wire command software, may cause an accident, with loss of life. As these systems are also very complex and are developed by large teams and maintained over long periods, their verification has become a crucial challenge. Safety critical systems are not limited to avionics: software runtime errors in cruise control management systems were recently blamed for causing *unintended acceleration* in certain Toyota models (the case was settled with a 1.2 billion dollars fine in March 2014, after years of investigation and several trials). Similarly, other transportation systems (railway), energy production systems (nuclear power plants, power grid management), medical systems (pacemakers, surgery and patient monitoring systems), and value transfers in decentralized systems (smart contracts), rely on complex software, which should be verified.

Beyond the field of embedded systems, other pieces of software may cause very significant harm in the case of bugs, as demonstrated by the Heartbleed security hole: due to a wrong protocol implementation, many websites could leak private information, over years.

An important example of semantic properties is the class of *safety* properties. A safety property typically specifies that some (undesirable) event will never occur, whatever the execution of the program that is considered. For instance, the absence of runtime error is a very important safety property. Other important classes of semantic properties include *liveness* properties (i.e., properties that specify that some desirable event will eventually occur) such as termination and *security* properties, such as the absence of information flows from private to public channels.

All these software semantic properties are *not decidable*, as can be shown by reduction to the halting problem. Therefore, there is no chance to develop any fully automatic technique able to decide, for any system, whether or not it satisfies some given semantic property.

The classic development techniques used in industry involve testing, which is not sound, as it only gives information about a usually limited test sample: even after successful test-based validation, situations that were untested may generate a problem. Furthermore, testing is costly in the long term, as it should be re-done whenever the system to verify is modified. Machine-assisted verification is another approach which verifies human specified properties. However, this approach also presents a very significant cost, as the annotations required to verify large industrial applications would be huge.

By contrast, the **antique** group focuses on the design of semantic analysis techniques that should be *sound* (i.e., compute semantic properties that are satisfied by all executions) and *automatic* (i.e., with no human interaction), although generally *incomplete* (i.e., not able to compute the best —in the sense of: most precise— semantic property). As a consequence of incompleteness, we may fail to verify a system that is actually correct. For instance, in the case of verification of absence of runtime error, the analysis may fail to validate a program, which is safe, and emit *false alarms* (that is reports that possibly dangerous operations were not proved safe), which need to be discharged manually. Even in this case, the analysis provides information about the alarm context, which may help disprove it manually or refine the analysis.

The methods developed by the **antique** group are not limited to the analysis of software. We also consider complex biological systems (such as models of signaling pathways, i.e. cascades of protein interactions, which enable signal communication among and within cells), described in higher level languages, and use abstraction techniques to reduce their combinatorial complexity and capture key properties so as to get a better insight in the underlying mechanisms of these systems.

# 3. Research Program

## 3.1. Semantics

Semantics plays a central role in verification since it always serves as a basis to express the properties of interest, that need to be verified, but also additional properties, required to prove the properties of interest, or which may make the design of static analysis easier.

For instance, if we aim for a static analysis that should prove the absence of runtime error in some class of programs, the concrete semantics should define properly what error states and non error states are, and how program executions step from a state to the next one. In the case of a language like C, this includes the behavior of floating point operations as defined in the IEEE 754 standard. When considering parallel programs, this includes a model of the scheduler, and a formalization of the memory model.

In addition to the properties that are required to express the proof of the property of interest, it may also be desirable that semantics describe program behaviors in a finer manner, so as to make static analyses easier to design. For instance, it is well known that, when a state property (such as the absence of runtime error) is valid, it can be established using only a state invariant (i.e., an invariant that ignores the order in which states are visited during program executions). Yet searching for trace invariants (i.e., that take into account some properties of program execution history) may make the static analysis significantly easier, as it will allow it to make finer case splits, directed by the history of program executions. To allow for such powerful static analyses, we often resort to a *non standard semantics*, which incorporates properties that would normally be left out of the concrete semantics.

## 3.2. Abstract interpretation and static analysis

Once a reference semantics has been fixed and a property of interest has been formalized, the definition of a static analysis requires the choice of an *abstraction*. The abstraction ties a set of *abstract predicates* to the concrete ones, which they denote. This relation is often expressed with a *concretization function* that maps each abstract element to the concrete property it stands for. Obviously, a well chosen abstraction should allow one to express the property of interest, as well as all the intermediate properties that are required in order to prove it (otherwise, the analysis would have no chance to achieve a successful verification). It should also lend

itself to an efficient implementation, with efficient data-structures and algorithms for the representation and the manipulation of abstract predicates. A great number of abstractions have been proposed for all kinds of concrete data types, yet the search for new abstractions is a very important topic in static analysis, so as to target novel kinds of properties, to design more efficient or more precise static analyses.

Once an abstraction is chosen, a set of *sound abstract transformers* can be derived from the concrete semantics and that account for individual program steps, in the abstract level and without forgetting any concrete behavior. A static analysis follows as a result of this step by step approximation of the concrete semantics, when the abstract transformers are all computable. This process defines an *abstract interpretation* [27]. The case of loops requires a bit more work as the concrete semantics typically relies on a fixpoint that may not be computable in finitely many iterations. To achieve a terminating analysis we then use *widening operators* [27], which over-approximate the concrete union and ensure termination.

A static analysis defined that way always terminates and produces sound over-approximations of the programs behaviors. Yet, these results may not be precise enough for verification. This is where the art of static analysis design comes into play through, among others:

- the use of more precise, yet still efficient enough abstract domains;
- the combination of application-specific abstract domains;
- the careful choice of abstract transformers and widening operators.

## 3.3. Applications of the notion of abstraction in semantics

In the previous subsections, we sketched the steps in the design of a static analyzer to infer some family of properties, which should be implementable, and efficient enough to succeed in verifying non trivial systems.

The same principles can be applied successfully to other goals. In particular, the abstract interpretation framework should be viewed as a very general tool to *compare different semantics*, not necessarily with the goal of deriving a static analyzer. Such comparisons may be used in order to prove two semantics equivalent (i.e., one is an abstraction of the other and vice versa), or that a first semantics is strictly more expressive than another one (i.e., the latter can be viewed an abstraction of the former, where the abstraction actually makes some information redundant, which cannot be recovered). A classical example of such comparison is the classification of semantics of transition systems [26], which provides a better understanding of program semantics in general. For instance, this approach can be applied to get a better understanding of the semantics of a programming language, but also to select which concrete semantics should be used as a foundation for a static analysis, or to prove the correctness of a program transformation, compilation or optimization.

## 3.4. From properties to explanations

In many application domains, we can go beyond the proof that a program satisfies its specification. Abstractions can also offer new perspectives to understand how complex behaviors of programs emerge from simpler computation steps. Abstractions can be used to find compact and readable representations of sets of traces, causal relations, and even proofs. For instance, abstractions may decipher how the collective behaviors of agents emerge from the orchestration of their individual ones in distributed systems (such as consensus protocols, models of signaling pathways). Another application is the assistance for the diagnostic of alarms of a static analyzer.

Complex systems and software have often times intricate behaviors, leading to executions that are hard to understand for programmers and also difficult to reason about with static analyzers. Shared memory and distributed systems are notorious for being hard to reason about due to the interleaving of actions performed by different processes and the non-determinism of the network that might lose, corrupt, or duplicate messages. Reduction theorems, e.g., Lipton's theorem, have been proposed to facilitate reasoning about concurrency, typically transforming a system into one with a coarse-grained semantics that usually increases the atomic sections. We investigate reduction theorems for distributed systems and ways to compute the coarse-grained counter part of a system automatically. Compared with shared memory concurrency, automated methods to

reason about distributed systems have been less investigated in the literature. We take a programming language approach based on high-level programming abstractions. We focus on partially-synchronous communication closed round-based models, introduced in the distributed algorithms community for its simpler proof arguments. The high-level language is compiled into a low-level (asynchronous) programming language. Conversely, systems defined under asynchronous programming paradigms are decompiled into the high-level programming abstractions. The correctness of the compilation/decompilation process is based on reduction theorems (in the spirit of Lipton and Elrad-Francez) that preserve safety and liveness properties.

In models of signaling pathways, collective behavior emerges from competition for common resources, separation of scales (time/concentration), non linear feedback loops, which are all consequences of mechanistic interactions between individual bio-molecules (e.g., proteins). While more and more details about mechanistic interactions are available in the literature, understanding the behavior of these models at the system level is far from easy. Causal analysis helps explaining how specific events of interest may occur. Model reduction techniques combine methods from different domains such as the analysis of information flow used in communication protocols, and tropicalization methods that comes from physics. The result is lower dimension systems that preserve the behavior of the initial system while focusing of the elements from which emerges the collective behavior of the system.

The abstraction of causal traces offer nice representation of scenarios that lead to expected or unexpected events. This is useful to understand the necessary steps in potential scenarios in signaling pathways; this is useful as well to understand the different steps of an intrusion in a protocol. Lastly, traces of computation of a static analyzer can themselves be abstracted, which provides assistance to classify true and false alarms. Abstracted traces are symbolic and compact representations of sets of counter-examples to the specification of a system which help one to either understand the origin of bugs, or to find that some information has been lost in the abstraction leading to false alarms.

# 4. Application Domains

## 4.1. Verification of safety critical embedded software

The verification of safety critical embedded software is a very important application domain for our group. First, this field requires a high confidence in software, as a bug may cause disastrous events. Thus, it offers an obvious opportunity for a strong impact. Second, such software usually have better specifications and a better design than many other families of software, hence are an easier target for developing new static analysis techniques (which can later be extended for more general, harder to cope with families of programs). This includes avionics, automotive and other transportation systems, medical systems ...

For instance, the verification of avionics systems represent a very high percentage of the cost of an airplane (about 30 % of the overall airplane design cost). The state of the art development processes mainly resort to testing in order to improve the quality of software. Depending on the level of criticality of a software (at the highest levels, any software failure would endanger the flight) a set of software requirements are checked with test suites. This approach is both costly (due to the sheer amount of testing that needs to be performed) and unsound (as errors may go unnoticed, if they do not arise on the test suite).

By contrast, static analysis can ensure higher software quality at a lower cost. Indeed, a static analyzer will catch all bugs of a certain kind. Moreover, a static analysis run typically lasts a few hours, and can be integrated in the development cycle in a seamless manner. For instance, ASTRÉE successfully verified the absence of runtime error in several families of safety critical fly-by-wire avionic software, in at most a day of computation, on standard hardware. Other kinds of synchronous embedded software have also been analyzed with good results.

In the future, we plan to greatly extend this work so as to verify *other families of embedded software* (such as communication, navigation and monitoring software) and *other families of properties* (such as security and liveness properties).

Embedded software in charge of communication, navigation, and monitoring typically relies on a *parallel* structure, where several threads are executed concurrently, and manage different features (input, output, user interface, internal computation, logging ...). This structure is also often found in automotive software. An even more complex case is that of *distributed* systems, where several separate computers are run in parallel and take care of several sub-tasks of a same feature, such as braking. Such a logical structure is not only more complex than the synchronous one, but it also introduces new risks and new families of errors (deadlocks, data-races...). Moreover, such less well designed, and more complex embedded software often utilizes more complex data-structures than synchronous programs (which typically only use arrays to store previous states) and may use dynamic memory allocation, or build dynamic structures inside static memory regions, which are actually even harder to verify than conventional dynamically allocated data structures. Complex data-structures also introduce new kinds of risks (the failure to maintain structural invariants may lead to runtime errors, non termination, or other software failures). To verify such programs, we will design additional abstract domains, and develop new static analysis techniques, in order to support the analysis of more complex programming language features such as parallel and concurrent programming with threads and manipulations of complex data structures. Due to their size and complexity, the verification of such families of embedded software is a major challenge for the research community.

Furthermore, embedded systems also give rise to novel security concerns. It is in particular the case for some aircraft-embedded computer systems, which communicate with the ground through untrusted communication media. Besides, the increasing demand for new capabilities, such as enhanced on-board connectivity, e.g. using mobile devices, together with the need for cost reduction, leads to more integrated and interconnected systems. For instance, modern aircrafts embed a large number of computer systems, from safety-critical cockpit avionics to passenger entertainment. Some systems meet both safety and security requirements. Despite thorough segregation of subsystems and networks, some shared communication resources raise the concern of possible intrusions. Because of the size of such systems, and considering that they are evolving entities, the only economically viable alternative is to perform automatic analyses. Such analyses of security and confidentiality properties have never been achieved on large-scale systems where security properties interact with other software properties, and even the mapping between high-level models of the systems and the large software base implementing them has never been done and represents a great challenge. Our goal is to prove empirically that the security of such large scale systems can be proved formally, thanks to the design of dedicated abstract interpreters.

The long term goal is to make static analysis more widely applicable to the verification of industrial software.

## 4.2. Static analysis of software components and libraries

An important goal of our work is to make static analysis techniques easier to apply to wider families of software. Then, in the longer term, we hope to be able to verify less critical, yet very commonly used pieces of software. Those are typically harder to analyze than critical software, as their development process tends to be less rigorous. In particular, we will target operating systems components and libraries. As of today, the verification of such programs is considered a major challenge to the static analysis community.

As an example, most programming languages offer Application Programming Interfaces (API) providing ready-to-use abstract data structures (e.g., sets, maps, stacks, queues, etc.). These APIs, are known under the name of containers or collections, and provide off-the-shelf libraries of high level operations, such as insertion, deletion and membership checks. These container libraries give software developers a way of abstracting from low-level implementation details related to memory management, such as dynamic allocation, deletion and pointer handling or concurrency aspects, such as thread synchronization. Libraries implementing data structures are important building bricks of a huge number of applications, therefore their verification is paramount. We are interested in developing static analysis techniques that will prove automatically the correctness of large audience libraries such as Glib and Threading Building Blocks.

## 4.3. Models of mechanistic interactions between proteins

Computer Science takes a more and more important role in the design and the understanding of biological systems such as signaling pathways, self assembly systems, DNA repair mechanisms. Biology has gathered large data-bases of facts about mechanistic interactions between proteins, but struggles to draw an overall picture of how these systems work as a whole. High level languages designed in Computer Science allow one to collect these interactions in integrative models, and provide formal definitions (i.e., semantics) for the behavior of these models. This way, modelers can encode their knowledge, following a bottom-up discipline, without simplifying *a priori* the models at the risk of damaging the key properties of the system. Yet, the systems that are obtained this way suffer from combinatorial explosion (in particular, in the number of different kinds of molecular components, which can arise at run-time), which prevents from a naive computation of their behavior.

We develop various analyses based on abstract interpretation, and tailored to different phases of the modeling process. We propose automatic static analyses in order to detect inconsistencies in the early phases of the modeling process. These analyses are similar to the analysis of classical safety properties of programs. They involve both forward and backward reachability analyses as well as causality analyses, and can be tuned at different levels of abstraction. We also develop automatic static analyses in order to identify key elements in the dynamics of these models. The results of these analyses are sent to another tool, which is used to automatically simplify models. The correctness of this simplification process is proved by the means of abstract interpretation: this ensures formally that the simplification preserves the quantitative properties that have been specified beforehand by the modeler. The whole pipeline is parameterized by a large choice of abstract domains which exploits different features of the high level description of models.

## 4.4. Consensus

Fault-tolerant distributed systems provide a dependable service on top of unreliable computers and networks. Famous examples are geo-replicated data-bases, distributed file systems, or blockchains. Fault-tolerant protocols replicate the system and ensure that all (unreliable) replicas are perceived from the outside as one single reliable machine. To give the illusion of a single reliable machine "consensus" protocols force replicas to agree on the "current state" before making this state visible to an outside observer. We are interested in (semi-)automatically proving the total correctness of consensus algorithms in the benign case (messages are lost or processes crash) or the Byzantine case (processes may lie about their current state). In order to do this, we first define new reduction theorems to simplify the behaviors of the system and, second, we introduce new static analysis methods to prove the total correctness of adequately simplified systems. We focus on static analysis based Satisfiability Modulo Theories (SMT) solvers which offers a good compromise between automation and expressiveness. Among our benchmarks are Paxos, PBFT (Practical Byzantine Fault-Tolerance), and blockchain algorithms (Red-Belly, Tendermint, Algorand). These are highly challenging benchmarks, with a lot of non-determinism coming from the interleaving semantics and from the adversarial environment in which correct processes execute, environment that can drop messages, corrupt them, etc. Moreover, these systems were originally designed for a few servers but today are deployed on networks with thousands of nodes. The "optimizations" for scalability can no longer be overlooked and must be considered as integral part of the algorithms, potentially leading to specifications weaker than the so much desired consensus.

## 4.5. Models of growth

In systems and synthetic biology (engineered systems) one would like study the environment of a given cellular process (such as signaling pathways mentioned earlier) and the ways in which that process interacts with different resources provided by the host. To do this, we have built coarse-grained models of cellular physiology which summarize fundamental processes (transcription, translation, transport, metabolism). such models describe global growth in mechanistic way and allow one to plug the model of one's process of interest into a simplified and yet realistic and reactive model of the process interaction with its immediate environment. A first ODE-based deterministic version of this model [30] explaining the famous bacterial growth laws and how the allocation of resources to different genomic sectors depends on the growth conditions- was published in 2015 and has already received nearly 150 citations. The model also allows one to bridge between population

genetic models which describe cells in terms of abstract features and fitness and intra-cellular models. For instance, we find that fastest growing strategies are not evolutionary stable in competitive experiments. We also find that vastly different energy storage strategies exist[16]. In a recent article[17] in *Nature Communications* we build a stochastic version of the above model. We predict the empirical size and doubling time distributions as a function of growth conditions. To be able to fit the parameters of the model to available single-cell data (note that the fitting constraints are far tighter than in the deterministic case), we introduce new techniques for the approximation of reaction-division systems which generalize continuous approximations of Langevin type commonly used for pure reaction systems. We also use cross-correlations to visualize causality and modes in noise propagation in the model (in a way reminiscent to abstract computational traces mentioned earlier). In other work, we show how to connect our new class of models to more traditional ones stemming from "flux balance analysis" by introducing an allocation vector which allows one to assign a formal growth rate to a class of reaction systems [25].

# 5. New Software and Platforms

## 5.1. APRON

SCIENTIFIC DESCRIPTION: The APRON library is intended to be a common interface to various underlying libraries/abstract domains and to provide additional services that can be implemented independently from the underlying library/abstract domain, as shown by the poster on the right (presented at the SAS 2007 conference. You may also look at:
FUNCTIONAL DESCRIPTION: The Apron library is dedicated to the static analysis of the numerical variables of a program by abstract interpretation. Its goal is threefold: provide ready-to-use numerical abstractions under a common API for analysis implementers, encourage the research in numerical abstract domains by providing a platform for integration and comparison of domains, and provide a teaching and demonstration tool to disseminate knowledge on abstract interpretation.

- Participants: Antoine Miné and Bertrand Jeannet
- Contact: Antoine Miné
- URL: http://apron.cri.ensmp.fr/library/

## 5.2. Astrée

*The AstréeA Static Analyzer of Asynchronous Software*
KEYWORDS: Static analysis - Static program analysis - Program verification - Software Verification - Abstraction
SCIENTIFIC DESCRIPTION: Astrée analyzes structured C programs, with complex memory usages, but without dynamic memory allocation nor recursion. This encompasses many embedded programs as found in earth transportation, nuclear energy, medical instrumentation, and aerospace applications, in particular synchronous control/command. The whole analysis process is entirely automatic.

Astrée discovers all runtime errors including:

undefined behaviors in the terms of the ANSI C99 norm of the C language (such as division by 0 or out of bounds array indexing),

any violation of the implementation-specific behavior as defined in the relevant Application Binary Interface (such as the size of integers and arithmetic overflows),

any potentially harmful or incorrect use of C violating optional user-defined programming guidelines (such as no modular arithmetic for integers, even though this might be the hardware choice),

failure of user-defined assertions.

FUNCTIONAL DESCRIPTION: Astrée analyzes structured C programs, with complex memory usages, but without dynamic memory allocation nor recursion. This encompasses many embedded programs as found in earth transportation, nuclear energy, medical instrumentation, and aerospace applications, in particular synchronous control/command. The whole analysis process is entirely automatic.

Astrée discovers all runtime errors including: - undefined behaviors in the terms of the ANSI C99 norm of the C language (such as division by 0 or out of bounds array indexing), - any violation of the implementation-specific behavior as defined in the relevant Application Binary Interface (such as the size of integers and arithmetic overflows), - any potentially harmful or incorrect use of C violating optional user-defined programming guidelines (such as no modular arithmetic for integers, even though this might be the hardware choice), - failure of user-defined assertions.

Astrée is a static analyzer for sequential programs based on abstract interpretation. The Astrée static analyzer aims at proving the absence of runtime errors in programs written in the C programming language.

- Participants: Antoine Miné, Jérôme Feret, Laurent Mauborgne, Patrick Cousot, Radhia Cousot and Xavier Rival
- Partners: CNRS - ENS Paris - AbsInt Angewandte Informatik GmbH
- Contact: Patrick Cousot
- URL: http://www.astree.ens.fr/

## 5.3. AstréeA

*The AstréeA Static Analyzer of Asynchronous Software*
KEYWORDS: Static analysis - Static program analysis
SCIENTIFIC DESCRIPTION: AstréeA analyzes C programs composed of a fixed set of threads that communicate through a shared memory and synchronization primitives (mutexes, FIFOs, blackboards, etc.), but without recursion nor dynamic creation of memory, threads nor synchronization objects. AstréeA assumes a real-time scheduler, where thread scheduling strictly obeys the fixed priority of threads. Our model follows the ARINC 653 OS specification used in embedded industrial aeronautic software. Additionally, AstréeA employs a weakly-consistent memory semantics to model memory accesses not protected by a mutex, in order to take into account soundly hardware and compiler-level program transformations (such as optimizations). AstréeA checks for the same run-time errors as Astrée , with the addition of data-races.
FUNCTIONAL DESCRIPTION: AstréeA is a static analyzer prototype for parallel software based on abstract interpretation. The AstréeA prototype is a fork of the Astrée static analyzer that adds support for analyzing parallel embedded C software.

- Participants: Antoine Miné, Jérôme Feret, Patrick Cousot, Radhia Cousot and Xavier Rival
- Partners: CNRS - ENS Paris - AbsInt Angewandte Informatik GmbH
- Contact: Patrick Cousot
- URL: http://www.astreea.ens.fr/

## 5.4. ClangML

KEYWORD: Compilation
FUNCTIONAL DESCRIPTION: ClangML is an OCaml binding with the Clang front-end of the LLVM compiler suite. Its goal is to provide an easy to use solution to parse a wide range of C programs, that can be called from static analysis tools implemented in OCaml, which allows to test them on existing programs written in C (or in other idioms derived from C) without having to redesign a front-end from scratch. ClangML features an interface to a large set of internal AST nodes of Clang , with an easy to use API. Currently, ClangML supports all C language AST nodes, as well as a large part of the C nodes related to C++ and Objective-C.

- Participants: Devin Mccoughlin, François Berenger and Pippijn Van Steenhoven
- Contact: Xavier Rival
- URL: https://github.com/Antique-team/clangml/tree/master/clang

## 5.5. FuncTion

SCIENTIFIC DESCRIPTION: FuncTion is based on an extension to liveness properties of the framework to analyze termination by abstract interpretation proposed by Patrick Cousot and Radhia Cousot. FuncTion infers ranking functions using piecewise-defined abstract domains. Several domains are available to partition the ranking function, including intervals, octagons, and polyhedra. Two domains are also available to represent the value of ranking functions: a domain of affine ranking functions, and a domain of ordinal-valued ranking functions (which allows handling programs with unbounded non-determinism).
FUNCTIONAL DESCRIPTION: FuncTion is a research prototype static analyzer to analyze the termination and functional liveness properties of programs. It accepts programs in a small non-deterministic imperative language. It is also parameterized by a property: either termination, or a recurrence or a guarantee property (according to the classification by Manna and Pnueli of program properties). It then performs a backward static analysis that automatically infers sufficient conditions at the beginning of the program so that all executions satisfying the conditions also satisfy the property.

- Participants: Antoine Miné and Caterina Urban
- Contact: Caterina Urban
- URL: http://www.di.ens.fr/~urban/FuncTion.html

## 5.6. HOO

*Heap Abstraction for Open Objects*
FUNCTIONAL DESCRIPTION: JSAna with HOO is a static analyzer for JavaScript programs. The primary component, HOO, which is designed to be reusable by itself, is an abstract domain for a dynamic language heap. A dynamic language heap consists of open, extensible objects linked together by pointers. Uniquely, HOO abstracts these extensible objects, where attribute/field names of objects may be unknown. Additionally, it contains features to keeping precise track of attribute name/value relationships as well as calling unknown functions through desynchronized separation.

As a library, HOO is useful for any dynamic language static analysis. It is designed to allow abstractions for values to be easily swapped out for different abstractions, allowing it to be used for a wide-range of dynamic languages outside of JavaScript.

- Participant: Arlen Cox
- Contact: Arlen Cox

## 5.7. MemCAD

*The MemCAD static analyzer*
KEYWORDS: Static analysis - Abstraction
FUNCTIONAL DESCRIPTION: MemCAD is a static analyzer that focuses on memory abstraction. It takes as input C programs, and computes invariants on the data structures manipulated by the programs. It can also verify memory safety. It comprises several memory abstract domains, including a flat representation, and two graph abstractions with summaries based on inductive definitions of data-structures, such as lists and trees and several combination operators for memory abstract domains (hierarchical abstraction, reduced product). The purpose of this construction is to offer a great flexibility in the memory abstraction, so as to either make very efficient static analyses of relatively simple programs, or still quite efficient static analyses of very involved pieces of code. The implementation consists of over 30 000 lines of ML code, and relies on the ClangML front-end. The current implementation comes with over 300 small size test cases that are used as regression tests.

- Participants: Antoine Toubhans, François Berenger, Huisong Li and Xavier Rival
- Contact: Xavier Rival
- URL: http://www.di.ens.fr/~rival/memcad.html

## 5.8. KAPPA

*A rule-based language for modeling interaction networks*
KEYWORDS: Systems Biology - Modeling - Static analysis - Simulation - Model reduction
SCIENTIFIC DESCRIPTION: OpenKappa is a collection of tools to build, debug and run models of biological pathways. It contains a compiler for the Kappa Language, a static analyzer (for debugging models), a simulator, a compression tool for causal traces, and a model reduction tool.
FUNCTIONAL DESCRIPTION: Kappa is provided with the following tools: - a compiler - a stochastic simulator - a static analyzer - a trace compression algorithm - an ODE generator.

RELEASE FUNCTIONAL DESCRIPTION: On line UI, Simulation is based on a new data-structure (see ESOP 2017 ), New abstract domains are available in the static analyzer (see SASB 2016), Local traces (see TCBB 2018), Reasoning on polymers (see SASB 2018).

- Participants: Jean Krivine, Jérôme Feret, Kim Quyen Ly, Pierre Boutillier, Russ Harmer, Vincent Danos and Walter Fontana
- Partners: ENS Lyon - Université Paris-Diderot - HARVARD Medical School
- Contact: Jérôme Feret
- URL: http://www.kappalanguage.org/

## 5.9. QUICr

FUNCTIONAL DESCRIPTION: QUICr is an OCaml library that implements a parametric abstract domain for sets. It is constructed as a functor that accepts any numeric abstract domain that can be adapted to the interface and produces an abstract domain for sets of numbers combined with numbers. It is relational, flexible, and tunable. It serves as a basis for future exploration of set abstraction.

- Participant: Arlen Cox
- Contact: Arlen Cox

## 5.10. LCertify

KEYWORD: Compilation
SCIENTIFIC DESCRIPTION: The compilation certification process is performed automatically, thanks to a prover designed specifically. The automatic proof is done at a level of abstraction which has been defined so that the result of the proof of equivalence is strong enough for the goals mentioned above and so that the proof obligations can be solved by efficient algorithms.
FUNCTIONAL DESCRIPTION: Abstract interpretation, Certified compilation, Static analysis, Translation validation, Verifier. The main goal of this software project is to make it possible to certify automatically the compilation of large safety critical software, by proving that the compiled code is correct with respect to the source code: When the proof succeeds, this guarantees semantic equivalence. Furthermore, this approach should allow to meet some domain specific software qualification criteria (such as those in DO-178 regulations for avionics software), since it allows proving that successive development levels are correct with respect to each other i.e., that they implement the same specification. Last, this technique also justifies the use of source level static analyses, even when an assembly level certification would be required, since it establishes separately that the source and the compiled code are equivalent.ntees that no compiler bug did cause incorrect code to be generated.

- Participant: Xavier Rival
- Partners: CNRS - ENS Paris
- Contact: Xavier Rival
- URL: http://www.di.ens.fr/~rival/lcertify.html

## 5.11. Zarith

FUNCTIONAL DESCRIPTION: Zarith is a small (10K lines) OCaml library that implements arithmetic and logical operations over arbitrary-precision integers. It is based on the GNU MP library to efficiently implement arithmetic over big integers. Special care has been taken to ensure the efficiency of the library also for small integers: small integers are represented as Caml unboxed integers and use a specific C code path. Moreover, optimized assembly versions of small integer operations are provided for a few common architectures.

Zarith is currently used in the Astrée analyzer to enable the sound analysis of programs featuring 64-bit (or larger) integers. It is also used in the Frama-C analyzer platform developed at CEA LIST and Inria Saclay.

- Participants: Antoine Miné, Pascal Cuoq and Xavier Leroy
- Contact: Antoine Miné
- URL: http://forge.ocamlcore.org/projects/zarith

# 6. New Results

## 6.1. A Theoretical Foundation of Sensitivity in an Abstract Interpretation Framework

**Participants:** Xavier Rival [correspondant], Sukyoung Ryu, Se-Won Kim.

In [14], we formalize a framework to design static analyses that make use of sensitivity, using the general notion of cardinal power abstraction.

Program analyses often utilize various forms of *sensitivity* such as context sensitivity, call-site sensitivity, and object sensitivity. These techniques all allow for more precise program analyses, that are able to compute more precise program invariants, and to verify stronger properties. Despite the fact that sensitivity techniques are now part of the standard toolkit of static analyses designers and implementers, no comprehensive frameworks allow the description of all common forms of sensitivity. As a consequence, the soundness proofs of static analysis tools involving sensitivity often rely on *ad hoc* formalization, which are not always carried out in an abstract interpretation framework. Moreover, this also means that opportunities to identify similarities between analysis techniques to better improve abstractions or to tune static analysis tools can easily be missed.

In this work, we formalize a framework for the description of *sensitivity in static analysis*. Our framework is based on a powerful abstract domain construction, and utilizes reduced cardinal power to tie basic abstract predicates to the properties analyses are sensitive to. We formalize this abstraction, and the main abstract operations that are needed to turn it into a generic abstract domain construction. We demonstrate that our approach can allow for a more precise description of program states, and that it can also describe a large set of sensitivity techniques, both when sensitivity criteria are static (known before the analysis) or dynamic (inferred as part of the analysis), and sensitive analysis tuning parameters. Last, we show that sensitivity techniques used in state of the art static analysis tools can be described in our framework.

## 6.2. Memory Abstraction

### *6.2.1. Abstraction of arrays based on non contiguous partitions*

**Participants:** Jiangchao Liu, Xavier Rival [correspondant].

In [15], we studied the verification of components of embedded programs that utilize arrays to store dynamically chained data-structures. Furthermore, this work constitutes a significant part of Jiangchao Liu's PhD Thesis ([10]).

User-space programs rely on memory allocation primitives when they need to construct dynamic structures such as lists or trees. However, low-level OS kernel services and embedded device drivers typically avoid resorting to an external memory allocator in such cases, and store structure elements in contiguous arrays instead. This programming pattern leads to very complex code, based on data-structures that can be viewed and accessed either as arrays or as chained dynamic structures. The code correctness then depends on intricate invariants mixing both aspects. We propose a static analysis that is able to verify such programs. It relies on the combination of abstractions of the allocator array and of the dynamic structures built inside it. This approach allows to integrate program reasoning steps inherent in the array and in the chained structure into a single abstract interpretation. We report on the successful verification of several embedded OS kernel services and drivers.

### *6.2.2. Semantic-Directed Clumping of Disjunctive Abstract States*

**Participants:** Huisong Li, Francois Berenger, Bor-Yuh Evan Chang, Xavier Rival [correspondant].

In [29], we studied the semantic directed clumping of disjunctive abstract states. Furthermore, this work constitutes a significant part of Huisong Li's PhD Thesis ([9]).

To infer complex structural invariants, Shape analyses rely on expressive families of logical properties. Many such analyses manipulate abstract memory states that consist of separating conjunctions of basic predicates describing atomic blocks or summaries. Moreover, they use finite disjunctions of abstract memory states in order to account for dissimilar shapes. Disjunctions should be kept small for the sake of scalability, though precision often requires to keep additional case splits. In this context, deciding when and how to merge case splits and to replace them with summaries is critical both for the precision and for the efficiency. Existing techniques use sets of syntactic rules, which are tedious to design and prone to failure. In this paper, we design a semantic criterion to clump abstract states based on their silhouette which applies not only to the conservative union of disjuncts, but also to the weakening of separating conjunction of memory predicates into inductive summaries. Our approach allows to define union and widening operators that aim at preserving the case splits that are required for the analysis to succeed. We implement this approach in the MemCAD analyzer, and evaluate it on real-world C codes from existing libraries, including programs dealing with doubly linked lists, red-black trees and AVL-trees.

## 6.3. Static Analysis of JavaScript Code

### *6.3.1. Weakly Sensitive Analysis for Unbounded Iteration over JavaScript Objects*

**Participants:** Yoonseok Ko, Xavier Rival [correspondant], Sukyoung Ryu.

In [28], we studied composite object abstraction for the analysis JavaScript.

JavaScript framework libraries like jQuery are widely use, but complicate program analyses. Indeed, they encode clean high-level constructions such as class inheritance via dynamic object copies and transformations that are harder to reason about. One common pattern used in them consists of loops that copy or transform part or all of the fields of an object. Such loops are challenging to analyze precisely, due to weak updates and as unrolling techniques do not always apply. In this work, we observe that precise field correspondence relations are required for client analyses (e.g., for call-graph construction), and propose abstractions of objects and program executions that allow to reason separately about the effect of distinct iterations without resorting to full unrolling. We formalize and implement an analysis based on this technique. We assess the performance and precision on the computation of call-graph information on examples from jQuery tutorials.

## 6.4. Communication-closed asynchronous protocols

**Participants:** Andrei Damien, Cezara Drăgoi [correspondant], Alexandru Militaru, Josef Widder.

Fault-tolerant distributed systems are implemented over asynchronous networks, so that they use algorithms for asynchronous models with faults. Due to asynchronous communication and the occurrence of faults (e.g., process crashes or the network dropping messages) the implementations are hard to understand and analyze.

In contrast, synchronous computation models simplify design and reasoning. In this paper, we bridge the gap between these two worlds. For a class of asynchronous protocols, we introduce a procedure that, given an asynchronous protocol, soundly computes its round-based synchronous counterpart. This class is defined by properties of the sequential code. We computed the synchronous counterpart of known consensus and leader election protocols, such as, Paxos, and Chandra and Toueg's consensus. Using Verifast we checked the sequential properties required by the rewriting. We verified the round-based synchronous counter-part of Multi-Paxos, and other algorithms, using existing deductive verification methods for synchronous protocols.

## 6.5. Borel Kernels and their Approximation, Categorically

**Participants:** Fredrik Dahlqvist, Alexandra Silva, Vicent Danos [correspondant], Ilias Garnier.

In [12] is introduced a categorical framework to study the exact and approximate semantics of probabilistic programs. We construct a dagger symmetric monoidal category of Borel kernels where the dagger-structure is given by Bayesian inversion. We show functorial bridges between this category and categories of Banach lattices which formalize the move from kernel-based semantics to predicate transformer (backward) or state transformer (forward) semantics. These bridges are related by natural transformations, and we show in particular that the Radon-Nikodym and Riesz representation theorems-two pillars of probability theory-define natural transformations. With the mathematical infrastructure in place, we present a generic and endogenous approach to approximating kernels on standard Borel spaces which exploits the involutive structure of our category of kernels. The approximation can be formulated in several equivalent ways by using the func-torial bridges and natural transformations described above. Finally, we show that for sensible discretization schemes, every Borel kernel can be approximated by kernels on finite spaces, and that these approximations converge for a natural choice of topology. We illustrate the theory by showing two examples of how approximation can effectively be used in practice: Bayesian inference and the Kleene * operation of ProbNetKAT.

## 6.6. Static analysis of rule-based models

Thanks to rule-based modeling languages, we can assemble large sets of mechanistic protein-protein interactions within integrated models. Our goal would be to understand how the behavior of these systems emerges from these low-level interactions. Yet this is a quite long term challenge and it is desirable to offer intermediary levels of abstraction, so as to get a better understanding of the models and to increase our confidence within our mechanistic assumptions. To this extend, static analysis can be used to derive various abstractions of the semantics, each of them offering new perspectives on the models.

### *6.6.1. Trace approximation*

**Participants:** Jérôme Feret [correspondant], Kim Quyên Lý.

In [13], we propose an abstract interpretation of the behavior of each protein, in isolation. Given a model written in Kappa, this abstraction computes for each kind of proteins a transition system that describes which conformations this protein may take and how a protein may pass from one conformation to another one. Then, we use simplicial complexes to abstract away the interleaving order of the transformations between conformations that commute. As a result, we get a compact summary of the potential behavior of each protein of the model.

### *6.6.2. Detection of polymer formation*

**Participants:** Pierre Boutillier, Aurélie Faure de Pebeyre, Jérôme Feret [correspondant].

Rule-based languages, such as Kappa and BNGL, allow for the description of very combinatorial models of interactions between proteins. A huge (when not infinite) number of different kinds of bio-molecular compounds may arise due to proteins with multiple binding and phosphorylation sites. Knowing beforehand whether a model may involve an infinite number of different kinds of bio-molecular compounds is crucial for the modeler. On the first hand, having an infinite number of kinds of bio-molecular compounds is sometimes a hint for modeling flaws: forgetting to specify the conflicts among binding rules is a common mistake. On the second hand, it impacts the choice of the semantics for the models (among stochastic, differential, hybrid).

In [22], we introduce a data-structure to abstract the potential unbounded polymers that may be formed in a rule-based model. This data-structure is a graph, the nodes and the edges of which are labeled with patterns. By construction, every potentially unbounded polymer is associated to at least one cycle in that graph. This data-structure has two main advantages. Firstly, as opposed to site-graphs, one can reason about cycles without enumerating them (by the means of Tarjan's algorithm for detecting strongly connected components). Secondly, this data-structures may be combined easily with information coming from additional reachability analysis: the edges that are labeled with an overlap that is proved unreachable in the model may be safely discarded.

### *6.6.3. The static analyzer KaSa*

**Participants:** Pierre Boutillier, Ferdinanda Camporesi, Jean Coquet, Jérôme Feret [correspondant], Kim Quyên Lý, Nathalie Théret, Pierre Vignet.

KaSa is a static analyzer for Kappa models. Its goal is two-fold. Firstly, KaSa assists the modeler by warning about potential issues in the model. Secondly, KaSa may provide useful properties to check that what is implemented is what the modeler has in mind and to provide a quick overview of the model for the people who have not written it. The cornerstone of KaSa is a fix-point engine which detects some patterns that may never occur whatever the evolution of the system may be. From this, many useful information may be collected KaSa warns about rules that may never be applied, about potential irreversible transformations of proteins (that may not be reverted even thanks to an arbitrary number of computation steps) and about the potential formation of unbounded molecular compounds. Lastly, KaSa detects potential influences (activation/inhibition relation) between rules.

In [21], we illustrate the main features of KaSa on a model of the extracellular activation of the transforming growth factor, TGF-b.

## 6.7. The Kappa platform for rule-based modeling

**Participants:** Pierre Boutillier, Mutaamba Maasha, Xing Li, Héctor Medina-Abarca, Jean Krivine, Jérôme Feret [correspondant], Ioana Cristescu, Angus Forbes, Walter Fontana.

In [11], we present an overview of the Kappa platform, an integrated suite of analysis and visualization techniques for building and interactively exploring rule-based models. The main components of the platform are the Kappa Simulator, the Kappa Static Analyzer and the Kappa Story Extractor. In addition to these components, we describe the Kappa User Interface, which includes a range of interactive visualization tools for rule-based models needed to make sense of the complexity of biological systems. We argue that, in this approach, modeling is akin to programming and can likewise benefit from an integrated development environment. Our platform is a step in this direction.

We discuss details about the computation and rendering of static, dynamic, and causal views of a model, which include the contact map (CM), snapshots at different resolutions, the dynamic influence network (DIN) and causal compression. We provide use cases illustrating how these concepts generate insight. Specifically, we show how the CM and snapshots provide information about systems capable of polymerization, such as Wnt signaling. A well-understood model of the KaiABC oscillator, translated into Kappa from the literature, is deployed to demonstrate the DIN and its use in understanding systems dynamics. Finally, we discuss how pathways might be discovered or recovered from a rule-based model by means of causal compression, as exemplified for early events in EGF signaling.

The Kappa platform is available via the project website at kappa-language.org. All components of the platform are open source and freely available through the authors' code repositories.

## 6.8. Conservative approximation of systems of differential equations

We design a tools-kit to reason and abstract the solutions of the systems of differential equations that are described in high-level languages. Our abstractions are conservative in the sense that they provided sound lower and upper bounds for the value of some observables of the system. Our approach consists, firstly,

in inferring structural equalities about combinations of variables and structural inequalities about the value of variable derivatives thanks to symbolic reasoning at the level of the languages and, then, in using these numerical constraints to infer two differential equations for the variables of interest — one for the lower bound and one for the upper bound.

We focus on the systems of equations that are described in Kappa. Our goal is to provide a unifying framework that can deal with heterogeneous kinds of abstractions, including truncation, time- and concentration-scale separations, flow-based reduction, symmetries-based reduction.

### *6.8.1. Approximation of models of polymers*

**Participants:** Ken Chanseau Saint-Germain, Jérôme Feret [correspondant].

We propose a systematic approach to approximate the behavior of models of polymers synthesis/degradation, described in Kappa. Our abstraction consists in focusing on the behavior of all the patterns of size less than a given parameter. We infer symbolic equalities and inequalities which intentionally may be understood as algebraic constructions over patterns, and extensionally as sound properties about the concentration of the bio-molecular species that contain these patterns. Then, we derive a system of equations describing the time evolution of a lower and an upper bounds for the concentration of each pattern of interest.

This work has been presented at VEMDP 2018 (Verification of Engineered Molecular Devices and Programs), in Oxford, 19th July 2018, and at the days "BIOS-IA" of the working group BIOSS, at Pasteur Institute, Paris, 18th December 2018.

### *6.8.2. Approximation based on time- and/or concentration-scale separation*

**Participants:** Andreea Beica, Jérôme Feret [correspondant].

In [20], we have designed and tested an approximation method for ODE models of biochemical reaction systems, in which the guarantees are our major requirement. Borrowing from tropical analysis techniques, we look at the dominance relations among terms of each species' ODE. These dominance relations can be exploited to simplify the original model, by neglecting the dominated terms. As the dominant subsystems can change during the system's dynamics, depending on which species dominate the others, several possible modes exist. Thus, simpler models consisting of only the dominant subsystems can be assembled into hybrid, piece-wise smooth models, which approximate the behavior of the initial system. By combining the detection of dominated terms with symbolic bounds propagation, we show how to approximate the original model by an assembly of simpler models, consisting in ordinary differential equations that provide time-dependent lower and upper bounds for the concentrations of the initial models species. The utility of our method is twofold. On the one hand, it provides a reduction heuristics that performs without any prior knowledge of the initial system's behavior (i.e., no simulation of the initial system is needed in order to reduce it). On the other hand, our method provides sound interval bounds for each species, and hence can serve to evaluate the faithfulness of tropicalization reduction heuristics for ODE models of biochemical reduction systems. The method is tested on several case studies.

## 6.9. Sources, propagation and consequences of stochasticity in cellular growth

**Participants:** Philipp Thomas, Guillaume Terradot, Vicent Danos [correspondant], Andrea Weiße.

Growth impacts a range of phenotypic responses. Identifying the sources of growth variation and their propagation across the cellular machinery can thus unravel mechanisms that underpin cell decisions.

In [17], we present a stochastic cell model linking gene expression, metabolism and replication to predict growth dynamics in single bacterial cells. Alongside we provide a theory to analyze stochastic chemical reactions coupled with cell divisions, enabling efficient parameter estimation, sensitivity analysis and hypothesis testing. The cell model recovers population-averaged data on growth-dependence of bacterial physiology and how growth variations in single cells change across conditions. We identify processes responsible for this variation and reconstruct the propagation of initial fluctuations to growth and other processes. Finally, we study drug-nutrient interactions and find that antibiotics can both enhance and suppress growth heterogeneity. Our results provide a predictive framework to integrate heterogeneous data and draw testable predictions with implications for antibiotic tolerance, evolutionary and synthetic biology.

## 6.10. Survival of the Fattest: Evolutionary Trade-offs in Cellular Resource Storage

**Participants:** Guillaume Terradot, Andreea Beica, Andrea Weiße, Vicent Danos [correspondant].

Cells derive resources from their environments and use them to fuel the bio-synthetic processes that determine cell growth. Depending on how responsive the bio-synthetic processes are to the availability of intracellular resources, cells can build up different levels of resource storage.

In [16], we use a recent mathematical model of the coarse-grained mechanisms that drive cellular growth to investigate the effects of cellular resource storage on growth. We show that, on the one hand, there is a cost associated with high levels of storage resulting from the loss of stored resources due to dilution. We further show that, on the other hand, high levels of storage can benefit cells in variable environments by increasing biomass production during transitions from one medium to another. Our results thus suggest that cells may face trade-offs in their maintenance of resource storage based on the frequency of environmental change.

## 6.11. A Genetic Circuit Compiler: Generating Combinatorial Genetic Circuits with Web Semantics and Inference

**Participants:** William Waites, Goksel Misirli, Matteo Cavaliere, Vicent Danos [correspondant].

A central strategy of synthetic biology is to understand the basic processes of living creatures through engineering organisms using the same building blocks. Biological machines described in terms of parts can be studied by computer simulation in any of several languages or robotically assembled in vitro. In [19] we present a language, the Genetic Circuit Description Language (GCDL) and a compiler, the Genetic Circuit Compiler (GCC). This language describes genetic circuits at a level of granularity appropriate both for automated assembly in the laboratory and deriving simulation code. The GCDL follows Semantic Web practice and the compiler makes novel use of the logical inference facilities that are therefore available. We present the GCDL and compiler structure as a study of a tool for generating $\kappa$-language simulations from semantic descriptions of genetic circuits.

## 6.12. An Information-Theoretic Measure for Patterning in Epithelial Tissues

**Participants:** William Waites, Matteo Cavaliere, Élise Cachat, Vicent Danos [correspondant], Jamie A. Davies.

In [18], we present path entropy, an information-theoretic measure that captures the notion of patterning due to phase separation in organic tissues. Recent work has demonstrated, both in silico and in vitro, that phase separation in epithelia can arise simply from the forces at play between cells with differing mechanical properties. These qualitative results give rise to numerous questions about how the degree of patterning relates to model parameters or underlying biophysical properties. Answering these questions requires a consistent and meaningful way of quantifying degree of patterning that we observe. We define a resolution-independent measure that is better suited than image-processing techniques for comparing cellular structures. We show how this measure can be usefully applied in a selection of scenarios from biological experiment and computer simulation, and argue for the establishment of a tissue-graph library to assist with parameter estimation for synthetic morphology.

# 7. Partnerships and Cooperations

## 7.1. National Initiatives

### *7.1.1. AnaStaSec*

Title: Static Analysis for Security Properties

Type: ANR générique 2014

Defi: Société de l'information et de la communication

Instrument: ANR grant

Duration: January 2015 - December 2018

Coordinator: Inria Paris-Rocquencourt (France)

Others partners: Airbus France (France), AMOSSYS (France), CEA LIST (France), Inria Rennes-Bretagne Atlantique (France), TrustInSoft (France)

Inria contact: Jérôme Feret

See also: http://www.di.ens.fr/ feret/anastasec/

Abstract: An emerging structure in our information processing-based society is the notion of trusted complex systems interacting via heterogeneous networks with an open, mostly untrusted world. This view characterises a wide variety of systems ranging from the information system of a company to the connected components of a private house, all of which have to be connected with the outside.

It is in particular the case for some aircraft-embedded computer systems, which communicate with the ground through untrusted communication media. Besides, the increasing demand for new capabilities, such as enhanced on-board connectivity, e.g. using mobile devices, together with the need for cost reduction, leads to more integrated and interconnected systems. For instance, modern aircrafts embed a large number of computer systems, from safety-critical cockpit avionics to passenger entertainment. Some systems meet both safety and security requirements. Despite thorough segregation of subsystems and networks, some shared communication resources raise the concern of possible intrusions.

Some techniques have been developed and still need to be investigated to ensure security and confidentiality properties of such systems. Moreover, most of them are model-based techniques operating only at architectural level and provide no guarantee on the actual implementations. However, most security incidents are due to attackers exploiting subtle implementation-level software vulnerabilities. Systems should therefore be analyzed at software level as well (i.e. source or executable code), in order to provide formal assurance that security properties indeed hold for real systems.

Because of the size of such systems, and considering that they are evolving entities, the only economically viable alternative is to perform automatic analyses. Such analyses of security and confidentiality properties have never been achieved on large-scale systems where security properties interact with other software properties, and even the mapping between high-level models of the systems and the large software base implementing them has never been done and represents a great challenge. The goal of this project is to develop the new concepts and technologies necessary to meet such a challenge.

The project ANASTASEC project will allow for the formal verification of security properties of software-intensive embedded systems, using automatic static analysis techniques at different levels of representation: models, source and binary codes. Among expected outcomes of the project will be a set of prototype tools, able to deal with realistic large systems and the elaboration of industrial security evaluation processes, based on static analysis.

### *7.1.2. REPAS*

The project REPAS, Reliable and Privacy-Aware Software Systems via Bisimulation Metrics (coordination Catuscia Palamidessi, Inria Saclay), aims at investigating quantitative notions and tools for proving program correctness and protecting privacy, focusing on bisimulation metrics, the natural extension of bisimulation on quantitative systems. A key application is to develop mechanisms to protect the privacy of users when their location traces are collected. Partners: Inria (Comete, Focus), ENS Cachan, ENS Lyon, University of Bologna.

### *7.1.3. SAFTA*

Title: SAFTA Static Analysis for Fault-Tolerant distributed Algorithms.

Type: ANR JCJC 2018

Duration: February 2018 - February 2022

Coordinator: Cezara Drăgoi, CR Inria

Abstract: Fault-tolerant distributed data structures are at the core distributed systems. Due to the multiple sources of non-determinism, their development is challenging. The project aims to increase the confidence we have in distributed implementations of data structures. We think that the difficulty does not only come from the algorithms but from the way we think about distributed systems. In this project we investigate partially synchronous communication-closed round based programming abstractions that reduce the number of interleavings, simplifying the reasoning about distributed systems and their proof arguments. We use partial synchrony to define reduction theorems from asynchronous semantics to partially synchronous ones, enabling the transfer of proofs from the synchronous world to the asynchronous one. Moreover, we define a domain specific language, that allows the programmer to focus on the algorithm task, it compiles into efficient asynchronous code, and it is equipped with automated verification engines.

### *7.1.4. TGFSYSBIO*

Title: Microenvironment and cancer: regulation of TGF-$\beta$ signaling

Type: ANR générique 2014

Defi: Société de l'information et de la communication

Instrument: Plan Cancer 2014-2019

Duration: December 2015 - November 2018

Coordinator: INSERM U1085-IRSET

Others partners: Inria Paris (France), Inria Rennes-Bretagne Atlantique (France),

Inria contact: Jérôme Feret

Abstract: Most cases of hepatocellular carcinoma (HCC) develop in cirrhosis resulting from chronic liver diseases and the Transforming Growth Factor $\beta$ (TGF-$\beta$) is widely regarded as both the major pro-fibrogenic agent and a critical inducer of tumor progression and invasion. Targeting the deleterious effects of TGF-$\beta$ without affecting its physiological role is the common goal of therapeutic strategies. However, identification of specific targets remains challenging because of the pleiotropic effects of TGF-$\beta$ linked to the complex nature of its extracellular activation and signaling networks.

Our project proposes a systemic approach aiming at to identifying the potential targets that regulate the shift from anti- to pro-oncogenic effects of TGF-$\beta$. To that purpose, we will combine a rule-based model (Kappa language) to describe extracellular TGF-beta activation and large-scale state-transition based (Cadbiom formalism) model for TGF-$\beta$-dependent intracellular signaling pathways. The multi-scale integrated model will be enriched with a large-scale analysis of liver tissues using shotgun proteomics to characterize protein networks from tumor microenvironment whose remodeling is responsible for extracellular activation of TGF-$\beta$. The trajectories and upstream regulators of the final model will be analyzed with symbolic model checking techniques and abstract interpretation combined with causality analysis. Candidates will be classified with semantic-based approaches and symbolic bi-clustering technics. All efforts must ultimately converge to experimental validations of hypotheses and we will use our hepatic cellular models (HCC cell lines and hepatic stellate cells) to screen inhibitors on the behaviors of TGF-$\beta$ signal.

The expected results are the first model of extracellular and intracellular TGF-$\beta$ system that might permit to analyze the behaviors of TGF-$\beta$ activity during the course of liver tumor progression and to identify new biomarkers and potential therapeutic targets.

### 7.1.5. VeriAMOS

Title: Verification of Abstract Machines for Operating Systems

Type: ANR générique 2018

Defi: Société de l'information et de la communication

Instrument: ANR grant

Duration: January 2019 - December 2022

Coordinator: Inria Paris (France)

Others partners: LIP6 (France), IRISA (France), UGA (France)

Inria contact: Xavier Rival

Abstract: Operating System (OS) programming is notoriously difficult and error prone. Moreover, OS bugs can have a serious impact on the functioning of computer systems. Yet, the verification of OSes is still mostly an open problem, and has only been done using user-assisted approaches that require a huge amount of human intervention. The VeriAMOS proposal relies on a novel approach to automatically and fully verifying OS services, that combines Domain Specific Languages (DSLs) and automatic static analysis. In this approach, DSLs provide language abstraction and let users express complex policies in high-level simple code. This code is later compiled into low level C code, to be executed on an abstract machine. Last, the automatic static analysis verifies structural and robustness properties on the abstract machine and generated code. We will apply this approach to the automatic, full verification of input/output schedulers for modern supports like SSDs.

## 7.2. European Initiatives

### 7.2.1. FP7 & H2020 Projects

Type: IDEAS

Defi:

Instrument: ERC Proof of Concept Grant 2018

Objectif: Static Analysis for the VErification of Spreadsheets

Duration: January 2019 - June 2020

Coordinator: Inria (France)

Partner: None

Inria contact: Xavier Rival

Abstract: Spreadsheet applications (such as Microsoft Excel + VBA) are heavily used in a wide range of application domains including engineering, finance, management, statistics and health. However, they do not ensure robustness properties, thus spreadsheet errors are common and potentially costly. According to estimates, the annual cost of spreadsheet errors is around 7 billion dollars. For instance, in 2013, a series of spreadsheet errors at JPMorgan incurred 6 billion dollars trading losses. Yet, expert reports estimate about 90 % of the spreadsheets contain errors. The MemCAD ERC StG project opened the way to novel formal analysis techniques for spreadsheet applications. We propose to leverage these results into a toolbox able to safely *verify*, *optimize* and *maintain* spreadsheets, so as to reduce the likelihood of spreadsheet disasters. This toolbox will be commercialized by the startup MATRIXLEAD.

## 7.3. International Research Visitors

### 7.3.1. Visits of International Scientists

#### 7.3.1.1. Internships

Jérôme Feret has supervised the M1 Internship of Aurélie Faure de Pebeyre (AIV Master) and the M2 Internship of Albin Salazar (AIV Master).

Xavier Rival is supervising M1 Internships of Guillaume Reboullet and of Luc Chabassier (M1 at DIENS).

Vincent Danos supervised interns Raja Ben Ali and Jaime Aujaud (L2 Epitech).

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. General Chair, Scientific Chair*

- Jérôme Feret is a guest member of the Steering Committee of the Conference on Computational Methods in Systems Biology (CMSB).
- Jérôme Feret is a member of the Steering Committee of the Workshop on Static Analysis and Systems Biology (SASB).
- Xavier Rival organized the 60th meeting of the IFIP Working Group 2.4 held in Dijon, in July 2018.
- Xavier Rival is a member of the Steering Committee of the Static Analysis Symposium (SAS).
- Xavier Rival is a member of the Steering Committee of the Workshop on Tools for Automatic Program Analysis (TAPAS).

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Chair of Conference Program Committees*

- Xavier Rival served as Chair of the Artifact Evaluation Committee of SAS 2018 (Static Analysis Symposium).

*8.1.2.2. Member of the Conference Program Committees*

- Jérôme Feret served as a Member of the Program Committee of LICS 2018 (Logic in Computer Science).
- Jérôme Feret served as a Member of the Program Committee of VEMDP 2018 (Verification of Engineered Molecular Devices and Programs).
- Jérôme Feret served as a Member of the Program Committee of SASB 2018 (Static Analysis and Systems Biology Workshop).
- Jérôme Feret served as a Member of the Program Committee of SAS 2018 (Static Analysis Symposium).
- Jérôme Feret served as a Member of the Program Committee of CMSB 2018 (Computational Methods in Systems Biology).
- Jérôme Feret served as a Member of the Program Committee of VMCAI 2019 (Verification, Model Checking, and Abstract Interpretation)
- Jérôme Feret is serving as a Member of the Program Committee of CIBCB 2019 (Computational Intelligence in Bioinformatics and Computational Biology).
- érôme Feret is serving as a Member of the Program Committee of HSB 2019 (Hybrid Systems and Biology).
- Jérôme Feret is serving as a Member of the Program Committee of CMSB 2019 (Computational Methods in Systems Biology).
- Xavier Rival served as a Member of the Program Committee of SAS 2018 (Static Analysis Symposium).
- Xavier Rival served as a Member of the Program Committee of Web Design, Analysis, Programming and Implementation of the WWW'18 Conference.

- Xavier Rival served as a Member of the Extended Review Committee of PLDI 2018 (Programming Languages Design and Implementation).
- Xavier Rival served as a Member of the Program Committee of APLAS 2018 (Asian Programming Languages And Systems Symposium).
- Xavier Rival is serving as a Member of the Committee of POPL 2020 (Principles of Programming Languages).
- Cezara Drăgoi served as a member of the Program Committee of Computer Aided Verification CAV'18.
- Cezara Drăgoi served as a member of the Program Committee of VMCAI 2019 (Verification, Model Checking, and Abstract Interpretation).
- Cezara Drăgoi served as a member of the Program Committee of NETYS 2018.

*8.1.2.3. Reviewer*

- Jérôme Feret served as Reviewer for ARSBM 2018 (Automated Reasoning for Systems Biology and Medicine).

### *8.1.3. Journal*

*8.1.3.1. Member of the Editorial Boards*

- Jérôme Feret serves as a Member of the Editorial Board of the Frontiers in Genetics journal and the Open Journal of Modeling and Simulation.
- Jérôme Feret serves as co-Editor of an Issue of the Theoretical Computer Science journal, that is composed of papers from SASB 2016, and is expected to appear in 2019.
- Jérôme Feret serves as co-Editor of an Issue of the IEEE/ACM Transactions on Computational Biology and Bioinformatics, that is composed of papers from CMSB 2017, and is expected to appear in 2019.
- Xavier Rival serves as Editor of an Issue of the Formal Methods in System Design Journal, that is composed of a selection of papers from SAS 2016, and appeared in 2018.

*8.1.3.2. Reviewer - Reviewing Activities*

- Jérôme Feret served as a Reviewer for NACO (Natural Computing).
- Jérôme Feret served as a Reviewer for ACS Synthetic Biology.
- Jérôme Feret served as a Reviewer for TCS (Theoretical Computer Sciences).
- Jérôme Feret served as a Reviewer for TCBB (IEEE/ACM Transactions on Computational Biology and Bioinformatics).
- Jérôme Feret served as a Reviewer for IEEE Transactions on Reliability.
- Xavier Rival served as a Reviewer for ACM TOPLAS (Transactions On Programming Languages and Systems).
- Xavier Rival served as a Reviewer for ACM TOPS (Transactions On Privacy and Security).

### *8.1.4. Invited Talks*

- Cezara Drăgoi was invited to give a talk at the workshop on Verification of Distributed Systems, Essaouira, Morocco, 2018.
- Cezara Drăgoi was invited to give a talk at the Dagstuhl workshop no 18211 on Formal Methods and Fault-Tolerant Distributed Computing: Forging an Alliance.

### *8.1.5. Leadership within the Scientific Community*

Xavier Rival is a member of the IFIP Working Group 2.4 on Software implementation technology.

### *8.1.6. Scientific Expertise*

- Cezara Drăgoi has participated to the recruitment committee for an assistant professor in the Department of Computer Science of École normale supérieure 2018.
- Jérôme Feret served as a external Reviewer for research program PRIM 2017 (funded by MIUR, the Italian Ministry for Education, University and Research).
- Jérôme Feret has participated to the recruitment committee for an assistant professor in Paris-Diderot University 2018.
- Xavier Rival chaired the Hiring Committee for an Assistant Professor position (Tenure track position, Gaspard Monge Chair) at École Polytechnique in 2018.

### *8.1.7. Research Administration*

- Jérôme Feret and Xavier Rival are members of the Laboratory Council of DIENS.
- Jérôme Feret is member of PhD Review Committee (CSD) of Inria Paris.
- Jérôme Feret is deputy head of study of the Department of Computer Science of École normale supérieure.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Licence:

- Marc Chevalier, Mathematics, 40h, L1, FDV Bachelor program (Frontiers in Life Sciences (FdV)), Université Paris-Descartes, France.
- Jérôme Feret and Xavier Rival (lectures), and Marc Chevalier (tutorials), “Semantics and Application to Verification”, 36h, L3, at École Normale Supérieure, France.
- Xavier Rival, “Introduction to Static Analysis”, 8h, L3, at École des Mines de Paris, France.
- Xavier Rival, “Programmation Avancée”, 18h, L3, at École Polytechnique, France.

Master:

- Xavier Rival, “Introduction to Static Analysis”, 24h, Internet of Things Master (retraining curriculum, EXED), France.
- Xavier Rival, “Protocol Safety and Verification”, 12h, M2, Advanced Communication Networks Master, France.
- Cezara Drăgoi, Jérôme Feret, Antoine Miné, and Xavier Rival, “Abstract Interpretation: application to verification and static analysis”, 72h, M2. Parisian Master of Research in Computer Science (MPRI), France.
- Vincent Danos and Jérôme Feret (with Jean Krivine), Computational Biology, 28h, M1. Interdisciplinary Approaches to Life Science (AIV), Master Program, Université Paris-Descartes, France.

Doctorat:

- Jérôme Feret, “Intrinsic Coarse-Graining of Models of Signalling pathways", 1h30, aDVANCES IN SYSTEMS AND SYNTHETIC BIOLOGY Modelling Complex Biological Systems in the Context of Genomics Thematic Research School, March 2018, Évry, France.

### *8.2.2. Supervision*

- PhD in progress: Marc Chevalier, Static analysis of Security Properties in Critical Embedded Software. started in 2017 and supervised by Jérôme Feret

- PhD in progress: Hugo Illous, Relational Shape Abstraction Based on Separation Logic, started in 2015 and supervised by Xavier Rival and Matthieu Lemerre (CEA)
- PhD in progress: Olivier Nicole, Verification of micro-kernels, started in 2018 and supervised by Xavier Rival and Matthieu Lemerre (CEA)
- PhD defended: Huisong Li, Disjunctive Shape Abstraction for Shared Data-Structures, started in 2014 and supervised by Xavier Rival
- PhD defended: Jiangchao Liu, Static Analysis for Numeric and Structural Properties of Array Contents, started in 2014 and supervised by Xavier Rival

#### *8.2.3. Juries*

- Xavier Rival served as a Reviewer in the Jury of the PhD of Ahmad Salim Al-Sibahi (Defense planned for the 11th of January, 2018).

## 8.3. Popularization

#### *8.3.1. Internal or external Inria responsibilities*

- Xavier Rival is member of the “Bureau du comité des projets” since October 2018.
- Xavier Rival is Chair of the Hiring Committee for Inria researchers at the center of Paris (CRCN) for 2019.

# 9. Bibliography

## Major publications by the team in recent years

[1] J. BERTRANE, P. COUSOT, R. COUSOT, J. FERET, L. MAUBORGNE, A. MINÉ, X. RIVAL.*Static Analysis and Verification of Aerospace Software by Abstract Interpretation*, in "Proceedings of the American Institute of Aeronautics and Astronautics (AIAA Infotech@Aerospace 2010)", Atlanta, Georgia, USA, American Institute of Aeronautics and Astronautics, 2010

[2] B. BLANCHET, P. COUSOT, R. COUSOT, J. FERET, L. MAUBORGNE, A. MINÉ, D. MONNIAUX, X. RIVAL.*A Static Analyzer for Large Safety-Critical Software*, in "Proceedings of the ACM SIGPLAN 2003 Conference on Programming Language Design and Implementation (PLDI'03)", ACM Press, June 7–14 2003, p. 196–207

[3] A. BOUAJJANI, C. DRAGOI, C. ENEA, M. SIGHIREANU.*On inter-procedural analysis of programs with lists and data*, in "Proceedings of the 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2011, San Jose, CA, USA, June 4-8, 2011", 2011, p. 578–589 [*DOI :* 10.1145/1993498.1993566], https://dl.acm.org/citation.cfm?id=1993498.1993566

[4] P. COUSOT.*Constructive Design of a Hierarchy of Semantics of a Transition System by Abstract Interpretation*, in "Theoretical Computer Science", 2002, vol. 277, n$^{o}$ 1–2, p. 47–103

[5] J. FERET, V. DANOS, J. KRIVINE, R. HARMER, W. FONTANA.*Internal coarse-graining of molecular systems*, in "Proceeding of the national academy of sciences", Apr 2009, vol. 106, n$^{o}$ 16

[6] L. MAUBORGNE, X. RIVAL.*Trace Partitioning in Abstract Interpretation Based Static Analyzers*, in "Proceedings of the 14th European Symposium on Programming (ESOP'05)", M. SAGIV (editor), Lecture Notes in Computer Science, Springer-Verlag, 2005, vol. 3444, p. 5–20

[7] A. MINÉ.*The Octagon Abstract Domain*, in "Higher-Order and Symbolic Computation", 2006, vol. 19, p. 31–100

[8] X. RIVAL.*Symbolic Transfer Functions-based Approaches to Certified Compilation*, in "Conference Record of the 31st Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", ACM Press, New York, United States, 2004, p. 1–13

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[9] H. LI.*Shape Abstractions with Support for Sharing and Disjunctions*, ENS Paris - Ecole Normale Supérieure de Paris ; PSL University, January 2018, https://hal.archives-ouvertes.fr/tel-01963082

[10] J. LIU.*Static Analysis on Numeric and Structural Properties of Array Contents*, ENS Paris ; PSL University, February 2018, https://hal.archives-ouvertes.fr/tel-01963108

## Articles in International Peer-Reviewed Journal

[11] P. BOUTILLIER, M. MAASHA, X. LI, H. F. MEDINA-ABARCA, J. KRIVINE, J. FERET, I. CRISTESCU, A. G. FORBES, W. FONTANA.*The Kappa platform for rule-based modeling*, in "Bioinformatics", July 2018, vol. 34, n$^{o}$ 13, p. i583-i592 [*DOI :* 10.1093/BIOINFORMATICS/BTY272], https://hal.inria.fr/hal-01962663

[12] F. DAHLQVIST, A. SILVA, V. DANOS, I. GARNIER.*Borel Kernels and their Approximation, Categorically*, in "Electronic Notes in Theoretical Computer Science", December 2018, vol. 341, p. 91-119, https://hal.archives-ouvertes.fr/hal-01976416

[13] J. FERET, K. Q. LÝ.*Local Traces: An Over-Approximation of the Behavior of the Proteins in Rule-Based Models*, in "IEEE/ACM Transactions on Computational Biology and Bioinformatics", July 2018, vol. 15, n$^{o}$ 4, p. 1124-1137, https://hal.inria.fr/hal-01967635

[14] S.-W. KIM, X. RIVAL, S. RYU.*A Theoretical Foundation of Sensitivity in an Abstract Interpretation Framework*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", August 2018, vol. 40, n$^{o}$ 3, p. 1-44, https://hal.archives-ouvertes.fr/hal-01963069

[15] J. LIU, L. CHEN, X. RIVAL.*Automatic Verification of Embedded System Code Manipulating Dynamic Structures Stored in Contiguous Regions*, in "IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems", November 2018, vol. 37, n$^{o}$ 11, p. 2311-2322, https://hal.archives-ouvertes.fr/hal-01963049

[16] G. TERRADOT, A. BEICA, A. Y. WEISSE, V. DANOS.*Survival of the Fattest: Evolutionary Trade-offs in Cellular Resource Storage*, in "Electronic Notes in Theoretical Computer Science", April 2018, vol. 335, p. 91-112 [*DOI :* 10.1016/J.ENTCS.2018.03.010], https://hal.archives-ouvertes.fr/hal-01976385

[17] P. THOMAS, G. TERRADOT, V. DANOS, A. Y. WEISSE.*Sources, propagation and consequences of stochasticity in cellular growth*, in "Nature Communications", December 2018, vol. 9, n$^{o}$ 1 [*DOI :* 10.1038/S41467-018-06912-9], https://hal.archives-ouvertes.fr/hal-01976406

[18] W. WAITES, M. CAVALIERE, E. CACHAT, V. DANOS, J. A. DAVIES.*An Information-Theoretic Measure for Patterning in Epithelial Tissues*, in "IEEE Access", 2018, vol. 6, p. 40302-40312 [*DOI :* 10.1109/ACCESS.2018.2853624], https://hal.archives-ouvertes.fr/hal-01976389

[19] W. WAITES, G. MISIRLI, M. CAVALIERE, V. DANOS, A. WIPAT.*A Genetic Circuit Compiler: Generating Combinatorial Genetic Circuits with Web Semantics and Inference*, in "ACS Synthetic Biology", November 2018, vol. 7, n$^{o}$ 12, p. 2812-2823, https://hal.archives-ouvertes.fr/hal-01985802

### International Conferences with Proceedings

[20] A. BEICA, J. FERET, T. PETROV.*Tropical Abstraction of Biochemical Reaction Networks with Guarantees*, in "SASB'18 - Static Analysis in Systems Biology, affiliated with Static Analysis Symposium", Freiburg, Germany, August 2018, https://hal.inria.fr/hal-01962674

[21] P. BOUTILLIER, F. CAMPORESI, J. COQUET, J. FERET, K. Q. LÝ, N. THÉRET, P. VIGNET.*KaSa: A Static Analyzer for Kappa*, in "CMSB 2018 - 16th International Conference on Computational Methods in Systems Biology", Brno, Czech Republic, M. ČEŠKA, D. ŠAFRÁNEK (editors), LNCS, Springer Verlag, September 2018, vol. 11095, p. 285-291 [*DOI :* 10.1007/978-3-319-99429-1_17], https://hal-univ-rennes1.archives-ouvertes.fr/hal-01888951

[22] P. BOUTILLIER, J. FERET, A. FAURE DE PEBEYRE.*Proving the absence of unbounded polymers in rule-based models*, in "Static Analysis and Systems Biology 2018", Freiburg im Breisgau, Germany, August 2018, https://hal.inria.fr/hal-01967632

[23] T. SUZANNE, A. MINÉ.*Relational Thread-Modular Abstract Interpretation Under Relaxed Memory Models*, in "APLAS 2018 - 16th Asian Symposium on Programming Languages and Systems", Wellington, New Zealand, Lecture Notes in Computer Science, December 2018, vol. 11275, p. 109-128 [*DOI :* 10.1007/978-3-030-02768-1_6], https://hal.inria.fr/hal-01953358

### Other Publications

[24] A. DAMIAN, C. DRAGOI, A. MILITARU, J. WIDDER.*Communication-closed asynchronous protocols*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01991415

## References in notes

[25] A. BEICA, V. DANOS.*Synchronous Balanced Analysis*, in "Proceedings of the International Workshop on Hybrid Systems Biology", Springer-Verlag, Berlin, Germany, September 2016, p. 85-94, https://hal.archives-ouvertes.fr/hal-01974696

[26] P. COUSOT.*Constructive design of a hierarchy of semantics of a transition system by abstract interpretation*, in "Electr. Notes Theor. Comput. Sci.", 1997, vol. 6, p. 77–102, http://dx.doi.org/10.1016/S1571-0661(05)80168-9

[27] P. COUSOT, R. COUSOT.*Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints*, in "Conference Record of the Fourth Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", ACM Press, New York, United States, 1977, p. 238–252

[28] Y. KO, X. RIVAL, S. RYU.*Weakly Sensitive Analysis for Unbounded Iteration over JavaScript Objects*, in "APLAS 2017 - 15th Asian Symposium on Programming Languages and Systems", Suzhou, China, LNCS, Springer, November 2017, vol. 10695, p. 148-168 [*DOI :* 10.1007/978-3-319-71237-6_8], https://hal.inria.fr/hal-01648680

[29] H. LI, F. BÉRENGER, B.-Y. E. CHANG, X. RIVAL.*Semantic-Directed Clumping of Disjunctive Abstract States* *, in "POPL 2017 - 44th ACM SIGPLAN Symposium on Principles of Programming Languages", Paris, France, ACM, January 2017, vol. 52, n$^{o}$ 1, p. 32-45 [*DOI :* 10.1145/3009837.3009881], https://hal.inria.fr/hal-01648679

[30] A. Y. WEISSE, D. A. OYARZUN, V. DANOS, P. S. SWAIN.*Mechanistic links between cellular trade-offs, gene expression, and growth*, in "Proceedings of the National Academy of Sciences of the United States of America ", March 2015, vol. 112, n$^{o}$ 9, p. E1038-E1047 [*DOI :* 10.1073/PNAS.1416533112], https://hal.archives-ouvertes.fr/hal-01976394

# Team AOSTE2

## Models and methods of analysis and optimization for systems with real-time and embedded contraints

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
**Paris**

THEME
**Embedded and Real-time Systems**

# Table of contents

# Team AOSTE2

*Creation of the Team: 2017 January 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.3. - Distributed Systems
A1.5.2. - Communicating systems
A2.1.1. - Semantics of programming languages
A2.1.9. - Synchronous languages
A2.1.10. - Domain-specific languages
A2.2.4. - Parallel architectures
A2.2.5. - Run-time systems
A2.3. - Embedded and cyber-physical systems
A2.3.1. - Embedded systems
A2.3.2. - Cyber-physical systems
A2.3.3. - Real-time systems
A2.4.1. - Analysis
A2.4.3. - Proofs
A8.2. - Optimization

**Other Research Topics and Application Domains:**

B5.2. - Design and manufacturing
B5.2.1. - Road vehicles
B5.2.2. - Railway
B5.2.3. - Aviation
B5.2.4. - Aerospace
B6.6. - Embedded systems

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Yves Sorel [Team leader, Inria, Senior Researcher]
Liliana Cucu [Inria, Researcher, HDR]
Robert Davis [University of York UK, Chair]
Dumitru Potop Butucaru [Inria, Researcher, HDR]

**External Collaborator**

Adriana Gogonel [StatInf, from September 2018]

**Technical Staff**

Adriana Gogonel [Inria, until Aug 2018]
Fatma Jebali [Inria]
Cristian Maxim [Inria, until Aug 2018]
Mehdi Mezouak [Inria]

**PhD Students**

Slim Ben Amor [Inria]

Keryan Didier [Inria]
Evariste Ntaryamira [Inria, Embassy of France at Burundi]
Salah Eddine Saidi [IFPEN, until Mar 2018]
Walid Talaboulma [Inria]

**Administrative Assistant**
Christine Anocq [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

The recent advances in merging different technologies and engineering domains has led to the emergence of Cyber-Physical Systems (CPS). In such systems, embedded computers interact with, and control physical processes. These embedded computers (cyber) may communicate from a tightly coupled way, for example through a serial CAN bus in the automotive domain or through an AFDX bus in the avionics domain to control engine(s) or brakes (physics), to a loosely coupled way for example through the internet network to offer multimedia services or data-base accesses. Because of the heterogeneity of the involved components (multi-physics, sensors, actuators, embedded computers), CPS may feature very complex design and implementation phases as well as complex computer platforms (multi/manycore, multiprocessor, distributed and parallel computers), ever raising the need for effective approaches in order to build reliable systems.

Most of these CPS are time sensitive, i.e. time is a crucial issue which must be carefully mastered, that yet increases their complexity. Mastering time in such CPS is the major objective of the team. Due to their heterogeneous nature, the different components may have different levels of criticality, e.g. engine and brakes have a higher criticality level than multimedia services, which increase the difficulty in the design and implementation phases since lower criticality parts must not interfere with higher criticality parts. In the team we mainly address mixed-criticality issues in term of software safety. However, we started to take into account, in addition, security issues (cyber attacks).

The members of the team beeing involved for a long time in *synchronous languages*, we address the functional specification of CPS with models compliant with the semantics of these languages. Theses models are basicaly graphs and more specifically “clocked graphs” that model data dependences beetween functions as well as “logical clocks” that are attached to every function. These logical clocks may be related to physical clocks which correspond to periods of functions. These periods are defined by automatic control engineers and are not dependent of the platform. Such approach allows verifications on the functional specification, guaranteeing that the output events of the control system obtained “in reaction” to some input events, are consistent with the input events that triggered them. Verifying functional specifications very early in the design phase, prevents a lot of classic errors found usually later on during the implementation phase. This approach is an important step for providing “correct by construction” implementations. However, non functional specifications must also be taken into consideration. Indeed, to perform real-time schedulability analyses used to guarantee that the implementation is correct in terms of time, we need the worst case execution times (WCET) of each function and the worst case communication times (WCCT) of each dependence. Both, worst case execution and communication times are dependent of the platform. Using these worst case times, schedulability analyses are able to compute the worst case response time of each function and some end-to-end worst case execution times, possibly in the case where the implied functions are allocated to different cores. Worst case response times and end-to-end execution times must verify real-time constraints, i.e. deadlines and latencies. These constraints are imposed by automatic control engineers while they usually do not know the platform that will be used later on in the developpement process.

This is the reason why, in the non functionnal specifications we need precise models that encompass important features found at different levels of the platform architecture, e.g. at a high level the number of cores and communications mediums, at a low level the structure of the caches, pipelines, etc. Depending on the complexity of the platform the problem of estimating these worst case times may be more or less difficult. In the case of simple predictable processors and buses, both used presently in the industry for critical avionics and railways applications, the estimation of worst case times is relatively easy. For this purpose we use, for example, static analyses or techniques based on measurements for estimating WCETs. However, due to the ever increasing smartphone market, the microprocessor industry provides more and more general purpose platforms based on multicore and, in a near future, based on manycore. These platform have complex architectures that are not predictable due to, e.g. multiple levels of cache and pipeline, speculative branching, communicating through shared memory or/and through a network on chip, etc. Therefore, nowadays the CPS industry has to face the great challenge of using such off the shelf platforms and consequently to estimate the corresponding worst case times of the programs (functions) that they will execute.

From functional and non functional specifications of the design phase we intend to synthesize, as automatically as possible, based on the real-time schedulability theory, an implementation that is correct by construction. This synthesizing process is close to the process used in language compilation but, in addition, it must take into account more complex non functional specifications. On the other hand, when platforms are not predictable an alternative to the classic estimation of worst case times mentioned previously, consists in reformulating the different problems in a probabilistic framework.

The overall objectives given above lead to three main research programs that are detailed below.

# 3. Research Program

## 3.1. The Algorithm-Architecture Adequation methodology and Real-Time Scheduling

**Participants:** Liliana Cucu, Dumitru Potop Butucaru, Yves Sorel.

The Algorithm-Architecture Adequation (AAA) methodology relies on distributed real-time schedulability and optimization theories to map efficiently an algorithm model to an architecture model.

The algorithm model which describes the functional specifications of the applications, is an extension of the well known data-flow model from Dennis [14]. It is a directed acyclic hyper-graph (DAG) that we call “conditioned factorized data dependence graph”, whose vertices are functions and hyper-edges are directed “data or control dependences” between functions. The data dependences define a partial order on the functions execution. The basic data-flow model was extended in three directions: first infinite (resp. finite) repetition of a sub-graph pattern in order to specify the reactive aspect of real-time systems (resp. in order to specify the finite repetition of a sub-graph consuming different data similar to a loop in imperative languages), second “state” when data dependences are necessary between different infinite repetitions of the sub-graph pattern introducing cycles which must be avoided by introducing specific vertices called “delays” (similar to z -n in automatic control), third “conditioning” of a function by a control dependence similar to conditional control structure in imperative languages, allowing the execution of alternative subgraphs. Delays combined with conditioning allow the programmer to specify automata used for describing “mode changes”.

The architecture model which describes the non functional specifications is, in the simplest case, a directed graph whose vertices are of two types: “processor” (one sequencer of functions, several sequencers of communications and distributed or shared memories) and “medium” (multiplexers and demultiplexers), and whose edges are directed connections. With such model it is possible to describe classic heterogeneous distributed, parallel and multiprocessor platforms as well as the most recent multi/manycore platforms. The worst case times mentioned previously are estimated according to this model.

The implementation model is a graph obtained by applying an external composition law such that an architecture graph operates on an algorithm graph to give an algorithm graph while taking advantage of timing characteristics, basically periods, deadlines and WCETs. This resulting algorithm graph is built by performing spatial and timing allocations (distribution and scheduling) of algorithm graph functions on architecture graph resources, and of dependences between functions on communication media. In that context "Adequation" means to search, in the solution space of implementation graphs, one implementation graph which verifies real-time constraints and, in addition, minimizes some criteria. These criteria consists in the total execution time of the algorithm executed on the architecture, the number of computing or communication resources, etc. Below, we describe distributed real-time schedulability analyses and optimization techniques suited for that purposes.

We address two main issues: uniprocessor and multiprocessor real-time scheduling for which some real-time constraints are of high criticality, i.e. they must be satisfied otherwise dramatic consequences occur.

In the case of uniprocessor real-time scheduling, besides the usual deadline constraint, often equal to the period of each task, i.e. a function with timing characteristics, we take into consideration dependences beetween tasks, and possibly several latencies. The latter are “end-to-end” constraints that may have complex relationships. Dealing with multiple real-time constraints raises the complexity of the scheduling problems. Moreover, costs of the Real-Time Operating System (RTOS) and of preemptions lead to, at least, a waste of resources due to their approximation in the WCET (Worst Execution Time) of each task, as proposed by Liu and Layland in their seminal article [21]. This is the reason why we first studied non-preemptive real-time scheduling with dependences, periodicities, and latencies constraints. Although a bad approximation of costs of the RTOS and of preemptions, may have dramatic consequences on real-time scheduling, there are only few researches on this topic. Thus, we investigated preemptive real-time scheduling while taking into account its cost which is very difficult to determine because it varies according to the instance (job) of each task. This latter is integrated in the schedulability conditions, and in the corresponding scheduling algorithms we propose. More generally, we integrate in schedulability analyses costs of the RTOS and of preemptions.

In the case of multiprocessor real-time scheduling, we chose to study first the “partitioned approach”, rather than the “global approach”, since the latter uses task migrations whose cost is prohibitive for current commercial processors, even for the more recent many/multicore. The partitioned approach enables us to reuse the results obtained in the uniprocessor case in order to derive solutions for the multiprocessor case. We consider also the semi-partitioned approach which allows only some migrations in order to minimize their costs. In addition, to satisfy the multiple real-time constraints mentioned in the uniprocessor case, we have to minimize the total execution time (makespan) since we deal with automatic control applications involving feedback loops. The complexity of such minimization problem increases because the cost of interprocessor communications (through buses in a multi-processor or routers in a manycore) must be taken into account. Furthermore, the domain of embedded systems leads to solving minimization resources problems. Since both optimization problems are NP-hard we develop exact algorithms (ILP, B & B, B & C) which are optimal for simple problems, and heuristics which are sub-optimal for realistic problems corresponding to industrial needs. Long time ago we proposed a very fast “greedy” heuristics whose results were regularly improved, and extended with local neighborhood heuristics, or used as initial solutions for metaheuristics.

Besides the spatial dimension (distributed) of the real-time scheduling problem, other important dimensions are the type of communication mechanisms (shared memory vs. message passing), or the source of control and synchronization (event-driven vs. time-triggered). We explore real-time scheduling on architectures corresponding to all combinations of the above dimensions. This is of particular impact in application domains such as railways and avionics.

## 3.2. Probabilistic Worst Case Reasoning for Real-Time Systems

**Participants:** Liliana Cucu, Robert Davis, Yves Sorel.

The arrival of modern hardware responding to the increasing demand for new functionalities exacerbates the limitations of the current worst-case real-time reasoning, mainly to the rarity of worst-case scenarios. Several

solutions exist to overcome this important pessimism and our solution takes into account the extremely low probability of appearance of a worst-case scenario within one hour of functioning ($10^{-45}$), compared to the certification requirements for instance ($10^{-9}$ for the highest level of certification in avionics). Thus we model and analyze real-time systems with time parameters described by using probabilistic models. Our results for such models address both schedulability analyses as well as timing analyses. Both such analyses are impacted by existing misunderstanding. The independence between tasks is a property of real-time systems that is often used for its basic results. Any complex model takes into account different dependences caused by sharing resources other than the processor. On another hand, the probabilistic operations require, generally, the (probabilistic) independence between the random variables describing some parameters of a probabilistic real-time system. The main (original) criticism to probabilistic is based on this hypothesis of independence judged too restrictive to model real-time systems. In reality the two notions of independence are different. Providing arguments to underline this confusion is at the center of our dissemination effort in the last years.

We provide below the bases driving our current research as follows:

- *Optimality of scheduling algorithms* stays an important aspect of the probabilistic real-time systems, especially that the introduction of probabilistic time parameters has a direct impact on the optimality of the existing scheduling algorithms. For instance Rate Monotonic scheduling policy is no longer optimal in the case of one processor when a preemptive fixed-priority solution exists. We expect other classes of algorithms to lose their optimality and we concentrate our efforts to propose new scheduling solutions in this context [22].
- *Increased complexity of schedulability analysis* due to the introduction of probabilistic parameters requires appropriate complexity reasoning, especially with the emergence of probabilistic schedulability analyses for mixed-criticality real-time systems [23]. Moreover the real-time applications are rarely independent and precedence constraint using graph-based models are appropriate in this context. Precedence constraints do decrease the number of possible schedulers, but they also imposes an "heritage" of probabilistic description from execution times to release times for instance.
- *Proving feasibility intervals* is crucial for these approaches that are often used in industry on top of simulation. As worst-case situations are rare events, then observing them or at least observe those events that do provoke later the appearance of worst-case situations is difficult. By proposing an iterative process of composition between different statistical models [17], we provide the basis to build a solution to this essential problem to prove any probabilistic real-time reasoning based on measurements.
- *Providing representativeness* of a measurement-based estimator is the final proof that a probabilistic worst-case reasoning may receive. Our first negative results [24] indicate that the measurement protocol is tighly connected to the statistical estimator and that both must verified properties of reproducibility in order to contribute to a convergence proof.

## 3.3. Real-Time Systems Compilation

**Participant:** Dumitru Potop Butucaru.

In the early days of embedded computing, most software development activities were manual. This is no longer true at the low level, where manual assembly coding has been almost completely replaced with the combined use of so-called “high-level” languages (C, Ada, *etc.*) and the use of compilers. This was made possible by the early adoption of standard interfaces that allowed the definition of economically-viable compilation tools with a large-enough user base. These interfaces include not only the programming languages (C, Ada, *etc.*), but also relatively stable microprocessor instruction set architectures (ISAs) or executable code formats like ELF.

The paradigm shift towards fully automated code generation is far from being completed at the system level, mainly due to the slower introduction of standard interfaces. This also explains why real-time scheduling has historically dedicated much of its research effort to verifying the correctness of very abstract and relatively standard implementation models (the task models). The actual construction of the implementations and the abstraction of these implementations as task models drew comparatively less interest, because they were application-dependent and non-portable.

But today the situation is bound to change. First, automation can no longer be avoided, as the complexity of systems steadily increases in both specification size (number of tasks, processors, etc.) and complexity of the objects involved (parallelized dependent tasks, multiple modes and criticalities, many-cores, *etc.*). Second, fully automated implementation is attainable for industrially significant classes of systems, due to significant advances in the standardization of both specification languages (Simulink, Scade, etc.) and of implementation platforms (ARINC, AUTOSAR, *etc.*).

To allow the automatic implementation of complex embedded systems, we advocate for a *real-time systems compilation* approach that combines aspects of both real-time scheduling – including the AAA methodology – and (classical) compilation. Like a classical compiler such as GCC, a real-time systems compiler should use fast and efficient scheduling and code generation heuristics, to ensure scalability. Similarly, it should provide traceability support under the form of informative error messages enabling an incremental trial-and-error design style, much like that of classical application software. This is more difficult than in a classical compiler, given the complexity of the transformation flow (creation of tasks, allocation, scheduling, synthesis of communication and synchronization code, *etc.*), and requires a full formal integration along the whole flow, including the crucial issue of correct hardware/platform abstraction.

A real-time systems compiler should perform precise, conservative timing accounting along the whole scheduling and code generation flow, allowing it to produce safe and tight real-time guarantees. In particular, resource allocation, timing analysis, and code generation must be tightly integrated to ensure that generated code (including communication and synchronization primitive calls) satisfies the timing hypotheses used for scheduling. More generally, and unlike in classical compilers, the allocation and scheduling algorithms must take into account a variety of non-functional requirements, such as real-time constraints, criticality/partitioning, preemptability, allocation constraints, *etc.* As the accent is put on the respect of requirements (as opposed to optimization of a metric, like in classical compilation), resulting scheduling problems are quite different from those of classical compilation.

We have designed and built a prototype real-time systems compiler, called LoPhT, for statically scheduled real-time systems. Results on industrial case studies are encouraging, hinting not only at the engineering potential of the approach, but also at the scientific research directions it opens.

One key issue here is sound hardware/platform abstraction. To prove that it is possible to reconcile performance with predictability in a fully automatic way, we started in the best possible configuration with industrial relevance: statically-scheduled software running on very predictable (yet realistic) platforms. Already, in this case, platform modeling is more complex than the one of classical compilation [0] or real-time scheduling. [0] The objective is now to move beyond this application class to more dynamic classes of specifications implementations, but without losing too much of the predictability and/or effciency.

Efficiency is also a critical issue in practical systems design, and we must invest more in the design of optimizations that improve the *worst-case* behavior of applications and take into account non-functional requirements in a *multi-objective optimization* perspective, but while remaining in the class of low-complexity heuristics to ensure scalability. Optimizations of classical compilation, such as loop unrolling, retiming, and inlining, can serve as inspiration.

Ensuring the safety and efficiency of the generated code cannot be done by a single team. Collaborations on the subject will have to cover at least the following subjects: the interaction between real-time scheduling and WCET analysis, the design of predictable hardware and software architectures, programming language support for efficient compilation, and formally proving the correctness of the compiler.

# 4. Application Domains

## 4.1. Avionics

[0]Because safe timing accounting is needed.

[0]The compiler must perform safe timing accounting, and not rely on experience-derived margins.

**Participants:** Liliana Cucu, Keryan Didier, Adriana Gogonel, Cristian Maxim, Dumitru Potop Butucaru, Yves Sorel.

A large number of our activities, in analysis, modelling, design and implementation of real-time embedded systems addresses specific applications mainly in the avionics field (with partners such as Airbus, Thales, Safran, etc.) (in the ASSUME project 9.2.1.1).

## 4.2. Many-Core Embedded Architectures

**Participants:** Liliana Cucu, Keryan Didier, Dumitru Potop Butucaru, Yves Sorel.

The AAA approach (fitting embedded applications onto embedded architectures) requires a sufficiently precise description of (a model of) the architecture (description platform). Such platforms become increasingly heterogeneous, and we had to consider a number of emerging ones with that goal in mind, such as Kalray MPPA (in the ASSUME project 9.2.1.1).

## 4.3. Railways

**Participants:** Liliana Cucu, Adriana Gogonel, Walid Talaboulma.

The statistical estimation of bounds on the execution time of a program on a processor is applied in the context of railroad crossing in the context of the collaborative project DEPARTS 9.1.2.2.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

This is the last activity report of the team AOSTE2 since it ends in 2018.

The ATT StatInf project, prepared by Liliana Cucu-Grosjean and Adriana Gogonel has been accepted in July 2018. The associated start-up creation has been selected for participation to the Digital Start-up program (jointly supported by EMLyon and Inria). The start-up will be created beginning of 2019 by Adriana Gogonel, Cristian Maxim and Liliana Cucu-Grosjean as founding members.

# 6. New Software and Platforms

## 6.1. SynDEx

KEYWORDS: Distributed - Optimization - Real time - Embedded systems - Scheduling analyses
SCIENTIFIC DESCRIPTION: SynDEx is a system level CAD software implementing the AAA methodology for rapid prototyping and for optimizing distributed real-time embedded applications. It is developed in OCaML.

Architectures are represented as graphical block diagrams composed of programmable (processors) and non-programmable (ASIC, FPGA) computing components, interconnected by communication media (shared memories, links and busses for message passing). In order to deal with heterogeneous architectures it may feature several components of the same kind but with different characteristics.

Two types of non-functional properties can be specified for each task of the algorithm graph. First, a period that does not depend on the hardware architecture. Second, real-time features that depend on the different types of hardware components, ranging amongst execution and data transfer time, memory, etc.. Requirements are generally constraints on deadline equal to period, latency between any pair of tasks in the algorithm graph, dependence between tasks, etc.

Exploration of alternative allocations of the algorithm onto the architecture may be performed manually and/or automatically. The latter is achieved by performing real-time multiprocessor schedulability analyses and optimization heuristics based on the minimization of temporal or resource criteria. For example while satisfying deadline and latency constraints they can minimize the total execution time (makespan) of the application onto the given architecture, as well as the amount of memory. The results of each exploration is visualized as timing diagrams simulating the distributed real-time implementation.

Finally, real-time distributed embedded code can be automatically generated for dedicated distributed real-time executives, possibly calling services of resident real-time operating systems such as Linux/RTAI or Osek for instance. These executives are deadlock-free, based on off-line scheduling policies. Dedicated executives induce minimal overhead, and are built from processor-dependent executive kernels. To this date, executives kernels are provided for: TMS320C40, PIC18F2680, i80386, MC68332, MPC555, i80C196 and Unix/Linux workstations. Executive kernels for other processors can be achieved at reasonable cost following these examples as patterns.
FUNCTIONAL DESCRIPTION: Software for optimising the implementation of embedded distributed real-time applications and generating efficient and correct by construction code
NEWS OF THE YEAR: We improved the distribution and scheduling heuristics to take into account the needs of co-simulation.

- Participant: Yves Sorel
- Contact: Yves Sorel
- URL: http://www.syndex.org

## 6.2. EVT Kopernic

KEYWORDS: Embedded systems - Worst Case Execution Time - Real-time application - Statistics
SCIENTIFIC DESCRIPTION: The EVT-Kopernic tool is an implementation of the Extreme Value Theory (EVT) for the problem of the statistical estimation of worst-case bounds for the execution time of a program on a processor. Our implementation uses the two versions of EVT - GEV and GPD - to propose two independent methods of estimation. Their results are compared and only results that are sufficiently close allow to validate an estimation. Our tool is proved predictable by its unique choice of block (GEV) and threshold (GPD) while proposant reproducible estimations.
FUNCTIONAL DESCRIPTION: EVT-Kopernic is tool proposing a statistical estimation for bounds on worst-case execution time of a program on a processor. The estimator takes into account dependences between execution times by learning from the history of execution, while dealing also with cases of small variability of the execution times.
NEWS OF THE YEAR: Any statistical estimator should come with an representative measurement protocole based on the processus of composition, proved correct. We propose the first such principle of composition while using a Bayesien modeling taking into account iteratively different measurement models. The composition model has been described in a patent submitted this year with a scientific publication under preparation.

- Participants: Adriana Gogonel and Liliana Cucu
- Contact: Adriana Gogonel
- URL: http://inria-rscript.serveftp.com/

## 6.3. LoPhT-manycore

*Logical to Physical Time compiler for many cores*
KEYWORDS: Real time - Compilation - Task scheduling - Automatic parallelization

SCIENTIFIC DESCRIPTION: Lopht is a system-level compiler for embedded systems, whose objective is to fully automate the implementation process for certain classes of embedded systems. Like in a classical compiler (e.g. gcc), its input is formed of two objects. The first is a program providing a platform-indepedent description of the functionality to implement and of the non-functional requirements it must satisfy (e.g. real-time, partitioning). This is provided under the form of a data-flow synchronous program annotated with non-functional requirements. The second is a description of the implementation platform, defining the topology of the platform, the capacity of its elements, and possibly platform-dependent requirements (e.g. allocation).

From these inputs, Lopht produces all the C code and configuration information needed to allow compilation and execution on the physical target platform. Implementations are correct by construction Resulting implementations are functionally correct and satisfy the non-functional requirements. Lopht-manycore is a version of Lopht targeting shared-memory many-core architectures.

The algorithmic core of Lopht-manycore is formed of timing analysis, allocation, scheduling, and code generation heuristics which rely on four fundamental choices. 1) A static (off-line) real-time scheduling approach where allocation and scheduling are represented using time tables (also known as scheduling or reservation tables). 2) Scalability, attained through the use of low-complexity heuristics for all synthesis and associated analysis steps. 3) Efficiency (of generated implementations) is attained through the use of precise representations of both functionality and the platform, which allow for fine-grain allocation of resources such as CPU, memory, and communication devices such as network-on-chip multiplexers. 4) Full automation, including that of the timing analysis phase.

The last point is characteristic to Lopht-manycore. Existing methods for schedulability analysis and real-time software synthesis assume the existence of a high-level timing characterization that hides much of the hardware complexity. For instance, a common hypothesis is that synchronization and interference costs are accounted for in the duration of computations. However, the high-level timing characterization is seldom (if ever) soundly derived from the properties of the platform and the program. In practice, large margins (e.g. 100%) with little formal justification are added to computation durations to account for hidden hardware complexity. Lopht-manycore overcomes this limitation. Starting from the worst-case execution time (WCET) estimations of computation operations and from a precise and safe timing model of the platform, it maintains a precise timing accounting throughout the mapping process. To do this, timing accounting must take into account all details of allocation, scheduling, and code generation, which in turn must satisfy specific hypotheses.
FUNCTIONAL DESCRIPTION: Accepted input languages for functional specifications include dialects of Lustre such as Heptagon and Scade v4. To ensure the respect of real-time requirements, Lopht-manycore pilots the use of the worst-case execution time (WCET) analysis tool (ait from AbsInt). By doing this, and by using a precise timing model for the platform, Lopht-manycore eliminates the need to adjust the WCET values through the addition of margins to the WCET values that are usually both large and without formal safety guarantees. The output of Lopht-manycore is formed of all the multi-threaded C code and configuration information needed to allow compilation, linking/loading, and real-time execution on the target platform.
NEWS OF THE YEAR: In the framework of the ITEA3 ASSUME project we have extended the Lopht-manycore to allow multiple cores to access the same memory bank at the same time. To do this, the timing accounting of Lopht has been extended to take into account memory access interferences during the allocation and scheduling process. Lopht now also pilots the aiT static WCET analysis tool from AbsInt by generating the analysis scripts, thus ensuring the consistency between the hypotheses made by Lopht and the way timing analysis is performed by aiT. As a result, we are now able to synthesize code for the computing clusters of the Kalray MPPA256 platform. Lopht-manycore is evaluated on avionics case studies in the perspective of increasing its technology readiness level for this application class.

- Participants: Dumitru Potop-Butucaru and Keryan Didier
- Contact: Dumitru Potop-Butucaru

# 7. New Results

## 7.1. Uniprocessor Mixed-Criticality Real-Time Scheduling

**Participants:** Liliana Cucu, Robert Davis, Mehdi Mezouak, Yves Sorel.

In the context of the FUI CEOS project 9.1.1.1, last year we tranformed the PX4 autopilot free software program in a graph of tasks. In this project our main goal is to perform a real-time schedulability analysis on this program in order to prove that the autopilot will meet all its deadlines when it will operate in the multirotor drone the project is intended to built. The tasks will be executed on a Pixhawk electronic board based on an ARM Cortex M4 microprocessor running on the NuttX OS.

We start by determinating the period and measuring the average execution time of each task which is less than the worst case execution time (WCET). Then, using these periods and these measured execution times we perform an online schedulability analysis using a rate monotonic policy (RM) that shown the set of tasks is not schedulable. Consequently, we informed the partners of the CEOS project that the present version of PX4 is not real-time.

Presently, we are transforming the original set of tasks into a set of real-time tasks. To achieve this goal, we associate to every task a periodic high resolution timer corresponding to the period of the task. Each timer generates an interruption when it expires and the task is put in the ready task queue. The scheduler of NuttX will choose in this queue the task to be executed. In order to validate this transformation we operated the multirotor drone in a simulation tool composed of Gazebo for the geometrical environment of the drone and of the Ground Control Station for setting and controlling the drone. We performed two kinds of simulations, a software in the loop simulation (SitL) which simulates the Pixhawk board, the sensors and the actuators, and a hardware in the loop simulation (HitL) which simulates only the sensors and the actuators, whereas the PX4 program runs on the Pixhawk board. We tested the set of real-time tasks in SitL and we are presently testing them in HitL.

Since we can easily change the period of every task, we plan to modify the periods to make the set of real-time tasks schedulable using an online RM schedulability analysis.

In order to manage high criticality real-time tasks we plan to use an offline scheduler whose scheduling table is generated by an offline schedulability analysis tool that is developped in the team. We plan to modify NuttX in order to support such scheduler.

Finally, in order to complete the real-time schedulability analysis of PX4, we estimate the worst case execution time (WCET) of each task. This problem is complex due to the multiple possible paths in a task as well as the different data it consumes. Moreover, the processor and/or the microcontroller itself may have some features like memory contentions, bus accesses, caches, pipelines, speculative branchings that increase the difficulty to determine WCETs. All these variabilities lead us to introduce statistical reasoning in characterizing the timing behavior (WCET, schedulability analyses) of mixed-criticality real-time applications. The isolated execution times of the programs have indicated large variations indicating expected larger variability in real execution scenarios. In order to decrease the pessimism of the statistical bounds, we are adapting our models to move towards multi-variate approaches.

## 7.2. Multiprocessor Real-Time Scheduling

**Participants:** Slim Ben Amor, Evariste Ntaryamira, Salah Eddine Saidi, Yves Sorel, Walid Talaboulma.

The last part of the PhD thesis of Salah Eddine Saidi, was dedicated to the parallelization of FMI-based co-simulation under real-time constraints. More precisely we address HiL (Hardware in the Loop) co-simulation where a part of the co-simulation is replaced by its real counterpart which is physically available. The real and simulated parts have to exchange data during the execution of the co-simulation under real-time constraints. In other words, the inputs (resp. ouputs) of the real part are sampled periodically, sending (resp. receiving) data to (resp. from) the simulated part. Every periodic data exchange defines a set of real-time constraints

to be satisfied by the simulated part. We proposed a method for defining these real-time constraints and propagating them to all the data dependent functions that specify the co-simulation (simulated part). Starting from these constraints we have to schedule the FMI-based co-simulation on a multi-core. We propose an ILP-based algorithm as well as a heuristic that allow the execution of the co-simulation on a multi-core processor while ensuring the previously defined real-time constraints are respected [6]. The proposed heuristic is a list scheduling heuristic. It builds the multi-core schedule iteratively. At each iteration, a list of candidate functions is constructed. The heuristic computes the priority for each candidate function on every core and selects the core for the which the priority is maximized. The priority of a function is a dynamic priority as its computation depends on the partial scheduling solution that has already been computed.

All works achieved by Salah Eddine Saidi on the parallelization of FMI-based co-simulation of numerical models were presented in his PhD thesis defense and manuscript [1].

Avionics applications are based on the specification of "data chains". Every data chain is a sequence of periodic real-time communicating tasks that are processing the data from sensors up to actuators. Such data chain determines an order in which the tasks propagate data but not in which they are executed. Indeed, inter-task communication and scheduling are independent. We focus on the latency computation, considered as the time elapsed from getting the data from an input and processing it to an output of a data chain. We propose a method for the worst-case latency computation of data chains composed of periodic tasks and executed by a partitioned fixed-priority preemptive scheduler upon a multiprocessor platform [5].

The PhD thesis of Slim Ben Amor is dedicated to the study of multiprocessor scheduling of real-time systems in presence of precedence constraints. This year we have proposed new models [10] for dependent real-time task with probabilistic worst-case execution time (WCET) that are scheduled using a partitioned reasoning. We explore existing solutions from [15] as the closest problem to our dependent task scheduling on multiprocessor and we study their extension to probabilistic models. We conclude that the probabilistic extension would be very difficult with heavy computation since the deterministic solution is based on the resolution of complex ILP optimization problem. Then, we decide to build a new solution to the deterministic problem that should be simple to extend to probabilistic problem. The proposed solution [11] consists of calculating the response time of each sub-tasks in a given DAG task taking in consideration preemptions caused by higher priority sub-tasks executed on the same processor. Then, we evaluate the global response time of the whole graph layer by layer, which allows deciding the schedulability of the entire system.

During the third year of Walid Talaboulma PhD thesis, we continued exploring solutions to make the WCET (Worst Case Execution Time) estimation as independent as possible with respect to the memory accesses. WCET analysis done on a unicore processor (in isolation) is not sufficient when we run our tasks on a multicore processors, the problem of Co-runner interference arises due to contention in shared hardware. Our solution is based on the generation of programs memory access profile, that we obtain by running tasks on a cycle accurate System Simulator, with a precise cycle accurate model of DDRAM memory controller and a full model of memory hierarchy including caches and main memory devices, and we log every memory event that occurs inside the simulation. Our solution does not necessarily require modifications of software layer, or recompilation of task code. We use those profiles to account for co runners interference and add it to WCET value obtained in isolation, and by updating our schedule, we can also insert idle times at correct scheduling events to decrease the interference.

The PhD thesis of Evariste Ntaryamira is dedicated to the study of multiprocessor real-time systems while ensuring the data freshness. This year we have underlined the difficulty of this scheduling problem [13], [8] while proposing a model to include both time and data constraints. We explore existing solutions from [16] as the closest problem to our data-dependent scheduling problem. The case study associated to this thesis is jointly prepared with the members of the RITS Inria team.

## 7.3. Safe Parallelization of Hard Real-Time Avionics Software

**Participants:** Keryan Didier, Dumitru Potop Butucaru.

This work took place in the framework of the ITEA3 ASSUME project, which funds the PhD thesis of Keryan Didier, and in close collaboration with Inria PARKAS, Airbus, Safran Aircraft Engines, and Kalray.

The key difficulty of real-time scheduling is that timing analysis and resource allocation depend on each other. An exhaustive search for the optimal solution not being possible for complexity reasons, heuristic approaches are used to break this dependency cycle. Two such approaches are typical in real-time systems design. The first approach uses unsafe timing characterizations for the tasks (e.g., measurements) to build the system, and then checks the respect of real-time requirements through a global timing analysis. The second approach uses a formal model of the hardware platform enabling timing characterizations that are safe for all possible resource allocations (worst-case bounds).

So far, the practicality of the second approach had never been established. Automated real-time parallelization flows still relied on simplified hypotheses ignoring much of the timing behavior of concurrent tasks, communication and synchronization code. And even with such unsafe hypotheses, few studies and tools considered the—harmonic—multiperiodic task graphs of real-world control applications, and the problem of statically managing all their computational, memory, synchronization and communication resources.

This year, we presented the first demonstration of the feasibility of the second approach, showing good practical results for classes of real-world applications and multiprocessor execution platforms whose timing predictability allows keeping pessimism under control. This requires something that is missing in previous work: *the tight orchestration of* **all** *implementation phases*: WCET analysis, resource allocation, generation of *glue code* ensuring the sequencing of tasks on cores and the synchronization and memory coherency between the cores, compilation and linking of the resulting C code. This orchestration is conducted on very detailed timing model that considers both the tasks and the generated glue code, and which includes resource access interferences due to multi-core execution. While orchestration is our main contribution, it should not be understood as a mere combination of existing tools and algorithms. The whole point of our approach is to carefully coordinate every analysis, mapping and code generation phase to enable predictable execution and to keep pessimism under control. To this end, we contributed application normalization phase to facilitate timing analysis, an original code generation algorithm designed to provide mapping-independent worst-case execution time bounds, and new real-time scheduling algorithms capable of orchestrating memory allocation and scheduling.

Our flow scales to an avionics application comprising more than 5000 unique nodes, targeting the Kalray MPPA 256 many-core platform, selected for its timing predictability. First results are presented in the report [9].

## 7.4. Real-time Platform Modeling

**Participants:** Fatma Jebali, Dumitru Potop Butucaru.

This work took place in the framework of the ITEA3 ASSUME project, which funds the post-doc of Fatma Jebali.

One key difficulty in embedded systems design is the existence of multiple models of the same hardware system, developed separately, at different abstraction levels, and used in various phases of the design flow. In the design of real-time embedded systems, we can identify, among other:

- Cycle-accurate system models used to perform fine-grain hardware simulation, mostly during HW and driver design phases. These models provide an exact functional and temporal representation of system execution.
- Microarchitectural models used for pipeline simulation during WCET (*Worst-Case Execution Time*) analysis [19], [20], [18]. These models are used to compute safe over-approximations of the duration of a sequential piece of code, i.e., one function running without interruption on a processor core). To provide precise results, these models preserve much of the microarchitectural detail of processor pipelines and memory hierarchy (e.g. cache states, data transfer latencies).

Both simulation models usually have cyclic activation patterns, but establishing semantic consistency between them is challenging for several reasons. First, the activation pattern, which is the logical time base of the simulation, depends on the abstraction level. In cycle-accurate models, simulation cycles correspond to hardware clock ticks, whereas in WCET analysis models they correspond to changes in the program counter of the sequential program. Second, data abstractions are different in the two simulation models. Cycle-accurate simulators are often also *bit-accurate*, *i.e.* provide exactly the same results as the actual hardware. By comparison, pipeline simulators in WCET analysis abstract away most data types and related operators, typically retaining only Booleans, which can be exploited at analysis time. Last, but not least, the simulators are usually pieces of C/C++ code manually written by different teams or obtained through complex translation processes from high-level Architecture Description Languages (ADLs) that may not have a clear semantics. Formally relating such pieces of code is difficult.

This year we proposed a method to ensure the semantic consistency between the two HW models we consider, focusing on time abstraction issues. Our method relies on *desynchronization* theory [25], which defines sufficient properties ensuring that a synchronous model can be seen as an asynchronous Kahn Process Network (KPN). When a synchronous HW model satisfies these properties, any scheduling of its computations that is compatible with data dependencies will produce the same result (a property known as scheduling-independence). We showed how to control scheduling through changes of the logical time base of the model prior to code generation using a synchronous language compiler. In particular, a careful choice of the logical time base allows us to produce, from the same model, either a cycle-accurate simulator, or the one needed for WCET analysis. In conjunction with some data abstraction, this logical time manipulation allows the synthesis of semantically consistent simulators from a single model.

Furthermore, we can ensure by construction that synchronous models satisfy the properties required by desynchronization theory. To this end, we introduced a new hardware modelling language, named xMAStime, allowing the compositional modeling of systems satisfying the required properties. Results were presented at the ACSD'18 conference [4].

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

The IFPEN grant which started on December 2014 and ended on February 2018, provides full support for the PhD thesis of Salah Eddine Saidi. The thesis concerns the automatic parallelization and scheduling approaches for co-simulation of numerical models on multi-core processors. The goal of the first research topic is to propose multi-core scheduling solutions for the co-simulation in order to accelerate its execution. The second research topic aims at proposing multi-core scheduling solutions in order to enable the execution of co-simulation under real-time constraints in the context of Hardware-in-the-Loop validation.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. FUI*

#### *9.1.1.1. CEOS*

**Participants:** Slim Ben Amor, Liliana Cucu, Cristian Maxim, Mehdi Mezouak, Yves Sorel, Walid Talaboulma.

This project was started on May 2017. Partners of the project are: ADCIS, ALERION, Aéroports de Lyon, EDF, ENEDIS, RTaW, EDF, Thales Communications and Security, ESIEE engineering school and Lorraine University. The CEOS project delivers a reliable and secure system of inspections of pieces of

works using professional mini-drone for Operators of Vital Importance coupled with their Geographical Information System. These inspections are carried out automatically at a lower cost than current solutions employing helicopters or off-road vehicles. Several software applications proposed by the industrial partners, are developed and integrated in the drone, within an innovative mixed-criticality approach using multi-core platforms.

*9.1.1.2. WARUNA*

**Participants:** Liliana Cucu, Adriana Gogonel, Yves Sorel, Walid Talaboulma.

This FUI funded project was started on September 2015 and it is preparing its final conclusions for the beginning of 2019. It has targeted the creation of the framework Time4Sys within the PolarSys project [12]. This open source framework allows timing analyses from models to simulation for different application domains like avionics, railways, medical, aerospace, automotive, etc. and it is available at https://www.polarsys.org/time4sys.

### *9.1.2. PIA*

*9.1.2.1. ES3CAP*

**Participants:** Keryan Didier, Dumitru Potop Butucaru.

The objectives of the ES3CAP (Embedded Smart Safe Secure Computing Autonomous Platform) project are to:

- Build a hardware and software industry-grade solution for the development of computation-intensive critical application. The solution should cover the needs of industrial end users, and target multi/many-core hardware platforms. The solution with come with 3 to 6 usage profiles specific to various industries (automotive, aerospace, defence).
- Improve the technology readiness level of the proposed development flow from TRL4-5 (technology development) to TRL6-7, thus approaching as much as possible commercialization.
- Build an alternate, perennial ecosystem for critical real-time OSs and development tools, for computer vision, data fusion and neural networks. The tools and components must be available on a prototyping and demonstration platform that is safe and secure.
- Capitalize on the convergence between the automotive and aerospace markets on subjects such as security, safety, decision making, and big data.

*9.1.2.2. DEPARTS*

**Participants:** Liliana Cucu, Adriana Gogonel, Walid Talaboulma.

This BGLE funded project of the national support programme Investissements d'Avenir has started on October 1st, 2012 and provided its final conclusions on December 2018. Inria has provided a final prototype version of the EVT Kopernic tool taking into account homogenous variation factors for the execution times. Swapping algorithms allowing WCET decrease are currently finalized within the PhD thesis of Walid Talaboulma with a defense expected during the spring of 2019.

## 9.2. European Initiatives

### *9.2.1. Collaborations in European Programs, Except FP7 & H2020*

*9.2.1.1. ASSUME*

**Participants:** Keryan Didier, Fatma Jebali, Dumitru Potop Butucaru.

Program: ITEA

Project acronym: ASSUME

Project title: Affordable Safe and Secure Mobility Evolution

Duration: September 2015 - August 2018

Coordinator: Daimler

Other partners: among 38 partners Absint, Ansys, Airbus, Kalray, Safran, Thales, ENS, KTH, FZI, etc.

Abstract: Future mobility solutions will increasingly rely on smart components that continuously monitor the environment and assume more and more responsibility for a convenient, safe and reliable operation. Currently the single most important roadblock for this market is the ability to come up with an affordable, safe multi-core development methodology that allows industry to deliver trustworthy new functions at competitive prices. ASSUME will provide a seamless engineering methodology, which addresses this roadblock on the constructive and analytic side.

### *9.2.2. Collaborations with Major European Organizations*

University of York: Real-Time System Group (UK)

Uncertainties in real-time systems: the utilization of extreme value theory has received increased efforts from our community and more rigorous principles are needed for its full understanding. Our two research teams have gathered these principles in several publications.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

Liliana Cucu-Grosjen is member of the steering committees of the following conferences and workshops: RTSS, RTAS, RTNS, WMC, RTSOPS.

Rob Davis is member of the steering committees of the following conferences and workshops: RTSS, RTAS, RTNS, WMC, RTSOPS.

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Chair of Conference Program Committees*

Liliana Cucu-Grosjean has served as PC co-chair for the 14th edition of WFCS 2018 in Imperia, Italy.

Adriana Gogonel has served as PC co-chair for the 12th Junior Researcher Workshop on Real-Time Computing (JWRTC) edition of in Poitiers, France.

*10.1.2.2. Member of the Conference Program Committees*

Liliana Cucu: RTAS, RTNS, WFCS.

Robert Davis: RTSS, RTAS, RTNS.

Adriana Gogonel: ACM RACS, WMC, JWRTC.

Dumitru Potop Butucaru: ACSD, EMSOFT.

Yves Sorel: DASIP.

*10.1.2.3. Reviewer*

All members of the team are regularly serving as reviewers for the main scientific events of our domain: RTSS, RTAS, RTCSA, RTNS, DATE, ETFA, EMSOFT, DASIP, etc.

### *10.1.3. Journal*

*10.1.3.1. Member of the Editorial Boards*

Liliana Cucu-Grosjean has served as guest editor for the Journal of Real-Time Systems

*10.1.3.2. Reviewer - Reviewing Activities*

All members of the team are regularly serving as reviewers for the main journals of our domain: Information Processing Letter, Journal of Heuristics, Journal of Real-Time Systems, Journal of Systems Architecture, Journal of Signal Processing Systems, Leibniz Transactions on Embedded Systems, IEEE Transactions on Industrial Informatics, IEEE Transactions on Computers, Theoretical Computer Science.

### *10.1.4. Scientific Expertise*

Yves Sorel: Steering Committee of System Design and Development Tools Group of Systematic Paris-Region Cluster.

Yves Sorel: Steering Committee of Technologies and Tools Program of SystemX Institute for Technological Research (IRT).

### *10.1.5. Research Administration*

Liliana Cucu-Grosjean is member of Inria Evaluation Commission, co-chair of Inria Committee on gender equality and equal oportunities, and member of the CLHCST.

Dumitru Potop Butucaru is member of mobility grant commission for postdocs and invited professors.

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master: Slim Ben Amor, Machine learning (practice sessions), 20H, M1, ESIEE Engineering School, Noisy-Le-Grand, France.

Master: Liliana Cucu, Distributed Databases and Statistics in Computer Science, 64h, U. Dunarea de Jos, Romania (Invited Professor).

Master: Liliana Cucu-Grosjean, Graph Theory, 32H, M1, ESIEE Engineering School, Cergy Pontoise, France.

Master: Adriana Gogonel, Machine learning, 32H, M1, ESIEE Engineering School, Noisy le Grand, France.

Master: Dumitru Potop Butucaru, A synchronous approach to the design of embedded real-time systems, 30h, M1, EPITA Engineering School, Paris France.

Master: Cristian Maxim, Graph Theory, 12H, M1, ESIEE Engineering School, Cergy Pontoise, France.

Master: Yves Sorel, Optimization of distributed real-time embedded systems, 38H, M2, University of Paris Sud, France.

Master: Yves Sorel, Safe design of reactive systems, 18H, M2, ESIEE Engineering School, Noisy-Le-Grand, France.

### *10.2.2. Supervision*

PhD: Salah-Edinne Saidi, Distributed real-time scheduling for the co-simulation of multiple control models, UPMC, defended April 2018, co-supervised by Nicolas Pernet (IFPEN) and Yves Sorel.

PhD in progress: Slim Ben Amor, Schedulability analysis of probabilistic real-time tasks under end to end constraints, UPMC, started on September 2016, supervised by Liliana Cucu.

PhD in progress: Keryan Didier, Formal certification of real-time implementations, UPMC, started November 2015, supervised by Dumitru Potop Butucaru.

PhD in progress: Evariste Ntaryamira, Analysis of embedded systems with time and security constraints, UPMC, started on January 2017, supervised by Liliana Cucu and Rachel Akimana.

PhD in progress: Walid Talaboulma, Probabilistic timing analysis in presence of dependences, UPMC, started November 2015, co-supervised by Liliana Cucu and Adriana Gogonel.

PhD in progress: Salah-Edinne Saidi, Distributed real-time scheduling for the co-simulation of multiple control models, UMPC, started December 2014, co-supervised by Nicolas Pernet (IFPEN) and Yves Sorel.

### *10.2.3. Juries*

Liliana Cucu-Grosjean is PhD examiner for the thesis of Anh Toan Bui Long, University of Poitiers/ENSMA, defended on December 2018.

Dumitru Potop Butucaru is PhD examiner for the thesis of Jad Khatib, Université Pierre et Marie Curie/EDITE, defended September 2018. École doctorale EDITE.

Yves Sorel is Phd examiner for the thesis of Florian Greff, University of Lorraine, defended May 2018.

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] S. E. SAIDI.*Automatic Parallelization and Scheduling Approches for Co-simulation of Numerical Models on Multi-core Processors*, Université Sorbonne, April 2018, https://hal.inria.fr/tel-01895280

### Articles in International Peer-Reviewed Journal

[2] L. CUCU-GROSJEAN, N. FISHER.*Guest editorial: special issue on real time and network systems*, in "Real-Time Systems", July 2018, vol. 54, n$^{o}$ 3, p. 605-606 [*DOI :* 10.1007/s11241-018-9309-8], https://hal.inria.fr/hal-01955995

[3] B. LESAGE, S. ALTMEYER, D. GRIFFIN, L. CUCU-GROSJEAN, R. DAVIS.*On the analysis of random replacement caches using static probabilistic timing methods for multi-path programs*, in "Real-Time Systems", April 2018, vol. 54, n$^{o}$ 2, p. 307-388 [*DOI :* 10.1007/s11241-017-9295-2], https://hal.archives-ouvertes.fr/hal-01666091

### International Conferences with Proceedings

[4] F. JEBALI, D. POTOP-BUTUCARU.*Ensuring consistency between cycle-accurate and instruction set simulators*, in "ACSD 2018 - 18th International Conference on Application of Concurrency to System Design", Bratislava, Slovakia, June 2018, https://hal.inria.fr/hal-01959370

[5] T. KLODA, A. BERTOUT, Y. SOREL.*Latency analysis for data chains of real-time periodic tasks*, in "ETFA'2018 - IEEE International Conference on Emerging Technologies and Factory Automation", Torino, Italy, Proceedings of the 23rd IEEE International Conference on Emerging Technologies and Factory Automation, ETFA'18, September 2018, https://hal.inria.fr/hal-01939228

[6] S. E. SAIDI, N. PERNET, Y. SOREL.*Scheduling Real-time HiL Co-simulation of Cyber-Physical Systems on Multi-core Architectures*, in "RTCSA2018 - IEEE International Conference on Embedded and Real-Time Computing Systems and Applications", Hakodate, Japan, Proceedings of the 24th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications, August 2018, https://hal.inria.fr/hal-01887155

[7] J. SOUYRIS, K. DIDIER, D. POTOP-BUTUCARU, G. IOOSS, T. BOURKE, A. COHEN, M. POUZET.*Automatic Parallelization from Lustre Models in Avionics*, in "ERTS2 2018 - 9th European Congress Embedded Real-Time Software and Systems", Toulouse, France, Proceedings of the 9th European Congress on Embedded Real-Time Software and Systems (ERTS2), 3AF - Association Aéronautique Astronautique de France and SEE - Société de l'électricité, de l'électronique et des technologies de l'information et de la communication and SIA - Société de Ingénieurs de l'Automobile, January 2018, p. 1-4, https://hal.inria.fr/hal-01714054

### Conferences without Proceedings

[8] E. NTARYAMIRA, C. MAXIM, L. CUCU-GROSJEAN.*Ensuring data freshness for periodic real-time tasks*, in "the 12th Junior Researcher Workshop on Real-Time Computing", Poitiers, France, 2018, https://hal.inria.fr/hal-01900328

### Research Reports

[9] K. DIDIER, D. POTOP-BUTUCARU, G. IOOSS, A. COHEN, J. SOUYRIS, P. BAUFRETON, A. GRAILLAT.*Parallelisation efficace de larges applications temps-reel*, Inria Paris, June 2018, n$^{o}$ RR-9180, https://hal.inria.fr/hal-01810176

### Other Publications

[10] S. BEN-AMOR, L. CUCU-GROSJEAN.*Probabilistic parallel real-time tasks model on multiprocessor platform*, July 2018, p. 1-2, RTSOPS 2018 - 9th International Real-Time Scheduling Open Problems Seminar, Poster, https://hal.inria.fr/hal-01956008

[11] S. BEN-AMOR, L. CUCU-GROSJEAN, D. MAXIM.*Response time analysis for precedence constrained and partitioned multiprocessor scheduled tasks*, October 2018, JWRTC 2018 - 12th Junior Researcher Workshop on Real-Time Computing, Poster, https://hal.inria.fr/hal-01957200

[12] L. FEJOZ, L. HAVET, A. DIDIER, B. VIAUD, A.-T. BUI LONG, T. D. NGUYEN, Y. OUHAMMOU, E. GROLLEAU, A. GOGONEL, C. MAXIM, L. CUCU-GROSJEAN, R. HENIA, L. RIOUX, N. SORDON, N. AYACHE, J. REHM.*Time4Sys – Integrating Timing Verification in your Engineering Practices*, December 2018, RTSS@Work 2018 - 39th IEEE Real-Time Systems Symposium Workshop, Poster, https://hal.inria.fr/hal-01957504

[13] E. NTARYAMIRA, C. MAXIM, C. FLORES, L. CUCU-GROSJEAN.*Towards temporal constraints in self driving cars*, July 2018, RTSOPS 2018 - 9th International Real-Time Scheduling Open Problems Seminar, Poster, https://hal.inria.fr/hal-01956016

## References in notes

[14] J. DENNIS.*First Version of a Dataflow Procedure Language*, in "Lecture Notes in Computer Sci.", Springer-Verlag, 1975, vol. 19, p. 362-376

[15] J. FONSECA, G. NELISSEN, V. NELIS, L. PINHO.*Response time analysis of sporadic DAG tasks under partitioned scheduling*, in "11th IEEE Symposium on Industrial Embedded Systems (SIES)", 05 2016, p. 1-10

[16] J. FORGET, E. GROLLEAU, C. PAGETTI, P. RICHARD.*Dynamic priority scheduling of periodic tasks with extended precedences*, in "IEEE 16th Conference on Emerging Technologies & Factory Automation, ETFA", 2011, p. 1–8

[17] A. GOGONEL, L. CUCU-GROSJEAN.*Dispositif de caractérisation et/ou de modélisation de temps d'exécution pire-cas*, June 2017, n$^{o}$ 1000408053, https://hal.archives-ouvertes.fr/hal-01666535

[18] D. HARDY, B. ROUXEL, I. PUAUT.*The Heptane Static Worst-Case Execution Time Estimation Tool*, in "Proceedings WCET", Dubrovnik, Croatia, 2017

[19] H. HERBEGUE, H. CASSÉ, M. FILALI, C. ROCHANGE.*Hardware architecture specification and constraint-based WCET computation*, in "Proceedings SIES", Porto, Portugal, 2013

[20] H. HERBEGUE, M. FILALI, H. CASSÉ.*Formal Architecture Specification for Time Analysis*, in "Proceedings ARCS", Lubeck, Germany, 2014

[21] C. LIU, J. LAYLAND.*Scheduling Algorithms for Multiprogramming in a Hard Real-Time Environment*, in "Journal of the ACM", January 1973, vol. 20, n$^{o}$ 1, p. 46-61

[22] D. MAXIM, L. CUCU-GROSJEAN, R. DAVIS.*Probabilistic schedulability analysis* , in "Handbook on Real-Time Computing", A. EASWARAN (editor), Handbook on Real-Time Computing, Springer, 2017, https://hal.archives-ouvertes.fr/hal-01666110

[23] D. I. MAXIM, R. DAVIS, L. CUCU-GROSJEAN, A. EASWARAN.*Probabilistic Analysis for Mixed Criticality Systems using Fixed Priority Preemptive Scheduling*, in "RTNS 2017 - International Conference on Real-Time Networks and Systems", Grenoble, France, October 2017, 10 [*DOI :* 10.1145/3139258.3139276], https://hal.inria.fr/hal-01614684

[24] C. MAXIM, A. GOGONEL, I. ASAVOAE, M. ASAVOAE, L. CUCU-GROSJEAN.*Reproducibility and representativity: mandatory properties for the compositionality of measurement-based WCET estimation approaches*, in "ACM SIGBED Review", November 2017, vol. 14, n$^{o}$ 3, p. 24-31 [*DOI :* 10.1145/3166227.3166230], https://hal.archives-ouvertes.fr/hal-01666084

[25] D. POTOP-BUTUCARU, B. CAILLAUD, A. BENVENISTE.*Concurrency in Synchronous Systems*, in "Formal Methods in System Design", Mar 2006, vol. 28, n$^{o}$ 2, p. 111–130

# Project-Team ARAMIS

## Algorithms, models and methods for images and signals of the human brain

IN COLLABORATION WITH: Institut du Cerveau et de la Moelle Epinière

IN PARTNERSHIP WITH:
**CNRS**

**INSERM**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Computational Neuroscience and Medicine**

# Table of contents

# Project-Team ARAMIS

*Creation of the Team: 2012 October 01, updated into Project-Team: 2014 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.2. - Unsupervised learning
A3.4.4. - Optimization and learning
A5.3. - Image processing and analysis
A5.4.4. - 3D and spatio-temporal reconstruction
A5.9. - Signal processing
A9. - Artificial intelligence
A9.2. - Machine learning
A9.3. - Signal analysis
A9.6. - Decision support

**Other Research Topics and Application Domains:**

B2. - Health
B2.2.6. - Neurodegenerative diseases
B2.6.1. - Brain imaging

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Olivier Colliot [CNRS, Team leader, Senior Researcher, HDR]
Ninon Burgos [CNRS, Researcher, from Oct 2018]
Fabrizio de Vico Fallani [Inria, Researcher, HDR]
Stanley Durrleman [Inria, Senior Researcher, HDR]

**Faculty Members**

Anne Bertrand [Sorbonne University, Associate Professor, deceased, March 2nd 2018]
Benjamin Charlier [Univ de Montpellier, Associate Professor, secondment at Inria]
Stephane Epelbaum [Assistance publique/Hôpitaux de Paris, Chair, Hospital Neurologist]
Didier Dormont [Sorbonne University, Professor]

**Technical Staff**

Simona Bottani [Institut du Cerveau et de la Moelle Epinière, until Sep 2018]
Mauricio Diaz Melo [Inria, from Apr 2018]
Clementine Fourrier [Institut du Cerveau et de la Moelle Epinière, until Mar 2018]
Juliana Gonzalez Astudillo [INSERM, from Oct 2018]
Arnaud Marcoux [Inria]
Benoit Martin [Inria, from May 2018]
Alexandre Routier [Inria, from Nov 2018]
Arnaud Valladier [Inria, from Nov 2018]

**PhD Students**

Manon Ansart [INSERM]
Giulia Bassignana [INSERM]

Simona Bottani [Inria, from Oct 2018]
Alexandre Bône [Univ Pierre et Marie Curie]
Tiziana Cattai [Inria]
Raphaël Couronné [Inria]
Johann Faouzi [Institut du Cerveau et de la Moelle Epinière]
Fanny Grosselin [MyBrainTechnologies]
Jeremy Guillon [Sorbonne University, until Sep 2018]
Igor Koval [INSERM]
Thomas Lartigue [Inria]
Maxime Louis [Sorbonne University]
Pascal Lu [Sorbonne University]
Federica Cacciamani [ICM, from Dec 2018]
Catalina Obando Forero [Inria, until Oct 2018]
Alexandre Routier [Sorbonne University, until Nov 2018]
Jorge Samper Gonzalez [Inria]
Elina Thibeau Sutre [Sorbonne University, from Oct 2018]
Wen Wei [Inria]
Junhao Wen [Sorbonne University]

**Post-Doctoral Fellows**

Ninon Burgos [Inria, until Sep 2018]
Alexis Guyot [Institut du Cerveau et de la Moelle Epinière]
Takoua Kaaouana [Institut du Cerveau et de la Moelle Epinière, until Mar 2018]
Vincent Henry [Inria, until Nov 2018]
Marie-Constance Corsi [Inria]

**Administrative Assistants**

Helene Milome [Inria]
Emmanuelle Mauduit [Institut du Cerveau et de la Moelle Epinière]

# 2. Overall Objectives

## 2.1. Context

**ARAMIS is an Inria project-team within the Brain and Spinal cord Institute (ICM - http://www.icm-institute.org) at the Pitié-Salpêtrière hospital in Paris**. ARAMIS was created as a team of the Inria Paris Center in 2012 and became a project-team in 2014. ARAMIS has a joint affiliation to Inria, CNRS, Inserm and Sorbonne University.

The **Pitié-Salpêtrière hospital** is the largest adult hospital in Europe. It is a leading center for neurological diseases: in terms of size (around 20,000 neurological patients each year), level of clinical expertise and quality of the technical facilities. Created in 2010, the **Brain and Spinal cord Institute (ICM)** gathers all research activities in neuroscience and neurology of the Pitié-Salpêtrière hospital. The ICM is both a private foundation and a public research unit (affiliated to CNRS, Inserm and University Pierre and Marie Curie). It hosts 25 research teams as well as various high level technical facilities (neuroimaging, genotyping/sequencing, cell culture, cellular imaging, bioinformatics ...), and gathers over 600 personnel. In addition, the ICM hosts one of the six IHU (*Instituts Hospitalo-Universitaires*), which are 10-year research programs funded for 55M euros each.

ARAMIS is thus located both within a leading neuroscience institute and within a large hospital. This unique position has several advantages: direct contact with neuroscientists and clinicians allows us to foresee the emergence of new problems and opportunities for new methodological developments, provides access to unique datasets, and eases the transfer of our results to clinical research and clinical practice.

## 2.2. General aim

The ARAMIS team is devoted to the design of **computational, mathematical and statistical approaches for the analysis of multimodal patient data**, with an emphasis on neuroimaging data. The core methodological domains of our team are: statistical and machine learning, statistical modeling of complex geometric data, connectivity and network analysis. These new approaches are applied to clinical research in neurological diseases in collaboration with other teams of the ICM, clinical departments of the Pitié-Salpêtrière hospital and external partners. **The team has a pluridisciplinary composition**, bringing together researchers in mathematics, computer science and engineering (N. Burgos, O. Colliot, F. De Vico Fallani, S. Durrleman) and clinicians (A. Bertrand, D. Dormont, S. Epelbaum). This general endeavor is addressed within the five following main objectives.

# 3. Research Program

## 3.1. From geometrical data to multimodal imaging

Brain diseases are associated to alterations of brain structure that can be studied in vivo using anatomical and diffusion MRI. The anatomy of a given subject can be represented by sets of anatomical surfaces (cortical and subcortical surfaces) and curves (white matter tracks) that can be extracted from anatomical and diffusion MRI respectively. We aim to develop approaches that can characterize the variability of brain anatomy within populations of subjects. To that purpose, we propose methods to estimate population atlases that provide an average model of a population of subjects together with a statistical model of their variability. Finally, we aim to introduce representations that can integrate geometrical information (anatomical surfaces, white matter tracts) together with functional (PET, ASL, EEG/MEG) and microstructural information.

## 3.2. Models of brain networks

Functional imaging techniques (EEG, MEG and fMRI) allow characterizing the statistical interactions between the activities of different brain areas, i.e. functional connectivity. Functional integration of spatially distributed brain regions is a well-known mechanism underlying various cognitive tasks, and is disrupted in brain disorders. Our team develops a framework for the characterization of brain connectivity patterns, based on connectivity descriptors from the theory of complex networks. More specifically, we propose analytical tools to infer brain networks, chacterize their structure and integrate multiple networks (for instance from multiple frequency bands or multiple modalities). The genericity of this approach allows us to apply it to various types of data including functional and structural neuroimaging, as well as genomic data.

## 3.3. Spatiotemporal modeling from longitudinal data

Longitudinal data sets are collected to capture variable temporal phenomena, which may be due to ageing or disease progression for instance. They consist in the observation of several individuals, each of them being observed at multiple points in time. The statistical exploitation of such data sets is notably difficult since data of each individual follow a different trajectory of changes and at its own pace. This difficulty is further increased if observations take the form of structured data like images or measurements distributed at the nodes of a mesh, and if the measurements themselves are normalized data or positive definite matrices for which usual linear operations are not defined. We aim to develop a theoretical and algorithmic framework for learning typical trajectories from longitudinal data sets. This framework is built on tools from Riemannian geometry to describe trajectories of changes for any kind of data and their variability within a group both in terms of the direction of the trajectories and pace.

## 3.4. Decision support systems

We then aim to develop tools to assist clinical decisions such as diagnosis, prognosis or inclusion in therapeutic trials. To that purpose, we leverage the tools developed by the team, such as multimodal representations, network indices and spatio-temporal models which are combined with advanced classification and regression approaches. We also dedicate strong efforts to rigorous, transparent and reproducible validation of the decision support systems on large clinical datasets.

## 3.5. Clinical research studies

Finally, we aim to apply advanced computational and statistical tools to clinical research studies. These studies are often performed in collaboration with other researchers of the ICM, clinicians of the Pitié -Salpêtrière hospital or external partners. Notably, our team is very often involved ”ex-ante” in clinical research studies. As co-investigators of such studies, we contribute to the definition of objectives, study design and definition of protocols. This is instrumental to perform clinically relevant methodological development and to maximize their medical impact. A large part of these clinical studies were in the field of dementia (Alzheimer's disease, fronto-temporal dementia). Recently, we expanded our scope to other neurodegenerative diseases (Parkinson's disease, multiple sclerosis).

# 4. Application Domains

## 4.1. Introduction

We develop different applications of our new methodologies to brain pathologies, mainly neurodegenerative diseases. These applications aim at:

- better understanding the pathophysiology of brain disorders;
- designing systems to support clinical decisions such as diagnosis, prognosis and design of clinical trials;
- developing brain computer interfaces for clinical applications.

## 4.2. Understanding brain disorders

Computational and statistical approaches have the potential to help understand the pathophysiology of brain disorders. We first aim to contribute to better understand the relationships between pathological processes, anatomical and functional alterations, and symptoms. Moreover, within a single disease, there is an important variability between patients. The models that we develop have the potential to identify more homogeneous disease subtypes, that would constitute more adequate targets for new treatments. Finally, we aim to establish the chronology of the different types of alterations. We focus these activities on neurodegeneratives diseases: dementia (Alzheimer's disease, fronto-temporal dementia), Parkinson's disease, multiple sclerosis.

## 4.3. Supporting clinical decisions

We aim to design computational tools to support clinical decisions, including diagnosis, prognosis and the design of clinical trials. The differential diagnosis of neurodegenerative diseases can be difficult. Our tools have the potential to help clinicians by providing automated classification that can integrate multiple types of data (clinical/cognitive tests, imaging, biomarkers). Predicting the evolution of disease in individual patients is even more difficult. We aim to develop approaches that can predict which alterations and symptoms will occur and when. Finally, new approaches are needed to select participants in clinical trials. Indeed, it is widely recognized that, to have a chance to be successful, treatments should be administered at a very early stage.

## 4.4. Brain computer interfaces for clinical applications

A brain computer interface (BCI) is a device aiming to decode brain activity, thus creating an alternate communication channel between a person and the external environment. BCI systems can be categorized on the basis of the classification of an induced or evoked brain activity. The central tenet of a BCI is the capability to distinguish different patterns of brain activity, each being associated to a particular intention or mental task. Hence adaptation, as well as learning, is a key component of a BCI because users must learn to modulate their brainwaves to generate distinct brain patterns. Usually, a BCI is considered a technology for people to substitute some lost functions. However, a BCI could also help in clinical rehabilitation to recover motor functions. Indeed, in current neuroscience-based rehabilitation it is recognized that protocols based on mental rehearsal of movements (like motor imagery practicing) are a way to access the motor system because they can induce an activation of sensorimotor networks that were affected by lesions. Hence, a BCI based on movement imagery can objectively monitor patients' progress and their compliance with the protocol, monitoring that they are actually imagining movements. It also follows that feedback from such a BCI can provide patients with an early reinforcement in the critical phase when there is not yet an overt sign of movement recovery.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

- The team has been awarded a "Fondation pour la Recherche sur la maladie d'Alzheimer" research grant.

### *5.1.1. Awards*

- Ninon Burgos received the Galileo Galilei Award 2017, best publication in the European Journal of Medical Physics - Physica Medica in 2017, for the paper 'Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning'.
- S. Durrleman successfully defended his "habilitation à diriger des Recherches" from Sorbonne University
- F. De Vico Fallani received the Young Investigator award from Complex Systems Society (CSS)
- Stéphane Epelbaum was awarded the Joel Ménard prize from the "Fondation Alzheimer".

# 6. New Software and Platforms

## 6.1. Brain Networks Toolbox

KEYWORDS: Neuroimaging - Medical imaging
FUNCTIONAL DESCRIPTION: Brain Networks Toolbox is an open-source package of documented routines implementing new graph algorithms for brain network analysis. It mainly contains Matlab code of new methods developed by the team and associated to publications (e.g., brain network thresholding, extraction of the information redundancy, node accessibility, etc). It requires, as input, adjacency matrices representing brain connectivity networks. Thus, it is independent on the specific approach used to construct brain networks and it can be used to extract network properties from any neuroimaging modality in healthy and diseased subjects.

- Participants: Fabrizio De Vico Fallani, Jeremy Guillon and Mario Chavez
- Contact: Fabrizio De Vico Fallani
- URL: https://github.com/brain-network/bnt

## 6.2. Deformetrica

KEYWORDS: 3D modeling - C++ - Automatic Learning - Mesh - Anatomy - Image analysis
SCIENTIFIC DESCRIPTION: Deformetrica is a software for the statistical analysis of 2D and 3D shape data. It essentially computes deformations of the 2D or 3D ambient space, which, in turn, warp any object embedded in this space, whether this object is a curve, a surface, a structured or unstructured set of points, or any combination of them.

Deformetrica comes with two applications:

registration, which computes the best possible deformation between two sets of objects, atlas construction, which computes an average object configuration from a collection of object sets, and the deformations from this average to each sample in the collection.

Deformetrica has very little requirements about the data it can deal with. In particular, it does not require point correspondence between objects!
FUNCTIONAL DESCRIPTION: Deformetrica is a software for the statistical analysis of 2D and 3D shape data. It essentially computes deformations of the 2D or 3D ambient space, which, in turn, warp any object embedded in this space, whether this object is a curve, a surface, a structured or unstructured set of points, or any combination of them.

Deformetrica comes with two applications:

- Registration, which computes the optimal deformation between two sets of objects,

- Atlas construction, which computes an average object configuration from a collection of object sets, and the deformations from this average to each sample in the collection.

Deformetrica has very little requirements about the data it can deal with. In particular, it does not require point correspondence between objects!

- Participants: Alexandre Routier, Ana Fouquier, Barbara Gris, Benjamin Charlier, Cédric Doucet, Joan Alexis Glaunès, Marcel Prastawa, Michael Bacci, Pietro Gori and Stanley Durrleman
- Partners: University of Utah - Université de Montpellier 2 - Université Paris-Descartes
- Contact: Stanley Durrleman
- URL: http://www.deformetrica.org/

## 6.3. Clinica

KEYWORDS: Neuroimaging - Brain MRI - MRI - Clinical analysis - Image analysis - Machine learning
SCIENTIFIC DESCRIPTION: Clinica is a software platform for multimodal brain image analysis in clinical research studies. It makes it easy to apply advanced analysis tools to large scale clinical studies. For that purpose, it integrates a comprehensive set of processing tools for the main neuroimaging modalities: currently MRI (anatomical, functional, diffusion) and PET, in the future, EEG/MEG. For each modality, Clinica allows to easily extract various types of features (regional measures, parametric maps, surfaces, curves, networks). Such features are then subsequently used as input of machine learning, statistical modeling, morphometry or network analysis methods. Processing pipelines are based on combinations of freely available tools developed by the community. It provides an integrated data management specification to store raw and processing data. Clinica is written in Python. It uses the Nipype system for pipelining. It combines widely-used software for neuroimaging data analysis (SPM, Freesurfer, FSL, MRtrix...), morphometry (Deformetrica), machine learning (Scikit-learn) and the BIDS standard for data organization.

FUNCTIONAL DESCRIPTION: Clinica is a software platform for multimodal brain image analysis in clinical research studies. It makes it easy to apply advanced analysis tools to large scale clinical studies. For that purpose, it integrates a comprehensive set of processing tools for the main neuroimaging modalities: currently MRI (anatomical, functional, diffusion) and PET, in the future, EEG/MEG. For each modality, Clinica allows to easily extract various types of features (regional measures, parametric maps, surfaces, curves, networks). Such features are then subsequently used as input of machine learning, statistical modeling, morphometry or network analysis methods. Clinica also provides an integrated data management specification to store raw and processing data. Overall, Clinica helps to: i) apply advanced analysis tools to clinical research studies, ii) easily share data and results, iii) make research more reproducible.
NEWS OF THE YEAR: - Three clinical studies made with Clinica Clinica : Bertrand et al, JAMA Neurology, 2018 , Jacquemont et al, Neurobiol Aging, 2017, Wen et al, JNNP, 2018 - Clinica presented at OHBM 2018 conference - Clinica was the support for the tutorial "Pattern Recognition for Neuroimaging" at OHBM 2018

- Participants: Jeremy Guillon, Thomas Jacquemont, Pascal Lu, Arnaud Marcoux, Tristan Moreau, Alexandre Routier, Jorge Samper Gonzalez, Junhao Wen, Olivier Colliot, Stanley Durrleman, Michael Bacci, Simona Bottani, Ninon Burgos, Sabrina Fontanella, Pietro Gori, Mauricio Diaz Melo and Elina Thibeau-Sutre
- Partners: Institut du Cerveau et de la Moelle épinière (ICM) - CNRS - INSERM - UPMC
- Contact: Olivier Colliot
- Publications: Amyloidosis and neurodegeneration result in distinct structural connectivity patterns in mild cognitive impairment - Yet Another ADNI Machine Learning Paper? Paving The Way Towards Fully-reproducible Research on Classification of Alzheimer's Disease - Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data - Neurite density is reduced in the presymptomatic phase of C9orf72 disease - Early cognitive, structural and microstructural changes in c9orf72 presymptomatic carriers before 40 years of age
- URL: http://www.clinica.run

## 6.4. Platforms

### *6.4.1. Platform Brain-computer interface*

Our team coordinates the developments of the Brain-Computer Interface (BCI) platform at the Centre EEG/MEG of the neuroimaging core facility of the ICM. Several projects, including our NETBCI NSF/NIH/ANR and ATTACK Big-brain theory funded projects, as well as experiments by different researchers of the Institute, are currently being run. To reinforce the impact of the platform we have recently recruited an engineer (J. Gonzalez-Astudillo) for the software and technical development.

# 7. New Results

## 7.1. Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data

**Participants:** Jorge Samper-González, Ninon Burgos, Simona Bottani, Sabrina Fontanella, Pascal Lu, Arnaud Marcoux, Alexandre Routier, Jérémy Guillon, Michael Bacci, Junhao Wen, Anne Bertrand, Hugo Bertin, Marie-Odile Habert, Stanley Durrleman, Theodoros Evgeniou, Olivier Colliot [Correspondant].

A large number of papers have introduced novel machine learning and feature extraction methods for automatic classification of Alzheimer's disease (AD). However, while the vast majority of these works use the public dataset ADNI for evaluation, they are difficult to reproduce because different key components of the validation are often not readily available. These components include selected participants and input data, image preprocessing and cross-validation procedures. The performance of the different approaches is also difficult to compare objectively. In particular, it is often difficult to assess which part of the method (e.g. preprocessing,

feature extraction or classification algorithms) provides a real improvement, if any. We proposed a framework for reproducible and objective classification experiments in AD using three publicly available datasets (ADNI, AIBL and OASIS). The framework comprises: i) automatic conversion of the three datasets into a standard format (BIDS); ii) a modular set of preprocessing pipelines, feature extraction and classification methods, together with an evaluation framework, that provide a baseline for benchmarking the different components. We demonstrated the use of the framework for a large-scale evaluation on 1960 participants using T1 MRI and FDG PET data. In this evaluation, we assessed the influence of different modalities, preprocessing, feature types (regional or voxel-based features), classifiers, training set sizes and datasets. Performances were in line with the state-of-the-art. FDG PET outperformed T1 MRI for all classification tasks. No difference in performance was found for the use of different atlases, image smoothing, partial volume correction of FDG PET images, or feature type. Linear SVM and L2-logistic regression resulted in similar performance and both outperformed random forests. The classification performance increased along with the number of subjects used for training. Classifiers trained on ADNI generalized well to AIBL and OASIS. All the code of the framework and the experiments is publicly available: general-purpose tools have been integrated into the Clinica software (http://www.clinica.run/) and the paper-specific code is available at: https://gitlab.icm-institute.org/aramislab/AD-ML.

More details in [30].

## 7.2. An automated pipeline for the analysis of PET data on the cortical surface

**Participants:** Arnaud Marcoux, Ninon Burgos, Anne Bertrand, Marc Teichmann, Alexandre Routier, Junhao Wen, Jorge Samper-González, Simona Bottani, Stanley Durrleman, Marie-Odile Habert, Olivier Colliot [Correspondant].

We developed a fully automatic pipeline for the analysis of PET data on the cortical surface. Our pipeline combines tools from FreeSurfer and PETPVC, and consists of i) co-registration of PET and T1-w MRI (T1) images, ii) intensity normalization, iii) partial volume correction, iv) robust projection of the PET signal onto the subject's cortical surface, v) spatial normalization to a template, and vi) atlas statistics. We evaluated the performance of the proposed workflow by performing group comparisons and showed that the approach was able to identify the areas of hypometabolism characteristic of different dementia syndromes: Alzheimer's disease (AD) and both the semantic and logopenic variants of primary progressive aphasia. We also showed that these results were comparable to those obtained with a standard volume-based approach. We then performed individual classifications and showed that vertices can be used as features to differentiate cognitively normal and AD subjects. This pipeline is integrated into Clinica, an open-source software platform for neuroscience studies available at http://www.clinica.run/.

More details in [24].

## 7.3. Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region

**Participants:** Hossein Arabi, Jason A. Dowling, Ninon Burgos [Correspondant], Xiao Han, Peter B. Greer, Nikolaos Koutsouvelis, Habib Zaidi.

Magnetic resonance imaging (MRI)-guided radiation therapy (RT) treatment planning is limited by the fact that the electron density distribution required for dose calculation is not readily provided by MR imaging. We compare a selection of novel synthetic CT generation algorithms recently reported in the literature, including segmentation-based, atlas-based and machine learning techniques, using the same cohort of patients and quantitative evaluation metrics. Six MRI-guided synthetic CT generation algorithms were evaluated: one segmentation technique into a single tissue class (water-only), four atlas-based techniques, namely, median value of atlas images (ALMedian), atlas-based local weighted voting (ALWV), bone enhanced atlas-based local weighted voting (ALWV-Bone), iterative atlas-based local weighted voting (ALWV-Iter), and a machine learning technique using deep convolution neural network (DCNN). Organ auto-contouring from MR images was evaluated for bladder, rectum, bones, and body boundary. Overall, DCNN exhibited higher segmentation

accuracy resulting in Dice indices while ALMedian showed the lowest accuracy. DCNN reached the best performance in terms of accurate derivation of synthetic CT values within each organ, followed by the advanced atlas-based methods. ALMedian led to the highest error. Considering the dosimetric evaluation results, ALWV-Iter, ALWV, DCNN and ALWV-Bone led to similar mean dose estimation within each organ at risk and target volume with less than 1% dose discrepancy. However, the two-dimensional gamma analysis demonstrated higher pass rates for ALWV-Bone, DCNN, ALMedian and ALWV-Iter at 1%/1 mm criterion. Overall, machine learning and advanced atlas-based methods exhibited promising performance by achieving reliable organ segmentation and synthetic CT generation. DCNN appears to have slightly better performance by achieving accurate automated organ segmentation and relatively small dosimetric errors (followed closely by advanced atlas-based methods, which in some cases achieved similar performance). However, the DCNN approach showed higher vulnerability to anatomical variation, where a greater number of outliers was observed with this method. Considering the dosimetric results obtained from the evaluated methods, the challenge of electron density estimation from MR images can be resolved with a clinically tolerable error.

More details in [4].

## 7.4. Double diffeomorphism: combining morphometry and structural connectivity analysis

**Participants:** Pietro Gori, Olivier Colliot, Linda Kacem, Yulia Worbe, Alexandre Routier, Cyril Poupon, Andreas Hartmann, Nicholas Ayache, Stanley Durrleman [Correspondant].

The brain is composed of several neural circuits which may be seen as anatomical complexes composed of grey matter structures interconnected by white matter tracts. Grey and white matter components may be modelled as 3D surfaces and curves respectively. Neurodevelopmental disorders involve morphological and organizational alterations which can not be jointly captured by usual shape analysis techniques based on single diffeomorphisms. We propose a new deformation scheme, called double diffeomorphism, which is a combination of two diffeomorphisms. The first one captures changes in structural connectivity, whereas the second one recovers the global morphological variations of both grey and white matter structures. This deformation model is integrated into a Bayesian framework for atlas construction. We evaluate it on a data-set of 3D structures representing the neural circuits of patients with Gilles de la Tourette syndrome (GTS). We show that this approach makes it possible to localise, quantify and easily visualise the pathological anomalies altering the morphology and organization of the neural circuits. Furthermore, results also indicate that the proposed deformation model better discriminates between controls and GTS patients than a single diffeomorphism.

More details in [15].

## 7.5. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms

**Participants:** Alexandre Bône, Olivier Colliot, Stanley Durrleman [Correspondant].

We propose a method to learn a distribution of shape tra-jectories from longitudinal data, i.e. the collection of individual objects repeatedly observed at multiple time-points. The method allows to compute an average spatiotemporal trajectory of shape changes at the group level, and the individual variations of this trajectory both in terms of geometry and time dynamics. First, we formulate a non-linear mixed-effects statistical model as the combination of a generic statistical model for manifold-valued longitudinal data, a deformation model defining shape trajectories via the action of a finite-dimensional set of diffeomorphisms with a man-ifold structure, and an efficient numerical scheme to compute parallel transport on this manifold. Second, we introduce a MCMC-SAEM algorithm with a specific approach to shape sampling, an adaptive scheme for proposal variances , and a log-likelihood tempering strategy to estimate our model. Third, we validate our algorithm on 2D simulated data, and then estimate a scenario of alteration of the shape of the hippocampus 3D brain structure during the course of Alzheimer's disease. The method shows for instance that hippocampal atrophy progresses more quickly in female subjects, and occurs earlier in APOE4 mutation carriers. We finally illustrate the potential of our method for classifying pathological trajectories versus normal ageing.

More details in [38].

## 7.6. Spatiotemporal Propagation of the Cortical Atrophy: Population and Individual Patterns

**Participants:** Igor Koval, Jean-Baptiste Schiratti, Alexandre Routier, Michael Bacci, Olivier Colliot, Stéphanie Allassonnière, Stanley Durrleman.

Repeated failures in clinical trials for Alzheimer's disease (AD) have raised a strong interest for the prodromal phase of the disease. A better understanding of the brain alterations during this early phase is crucial to diagnose patients sooner, to estimate an accurate disease stage, and to give a reliable prognosis. According to recent evidence, structural alterations in the brain are likely to be sensitive markers of the disease progression. Neuronal loss translates in specific spatiotemporal patterns of cortical atrophy, starting in the enthorinal cortex and spreading over other cortical regions according to specific propagation pathways. We developed a digital model of the cortical atrophy in the left hemisphere from prodromal to diseased phases, which is built on the temporal alignment and combination of several short-term observation data to reconstruct the long-term history of the disease. The model not only provides a description of the spatiotemporal patterns of cortical atrophy at the group level but also shows the variability of these patterns at the individual level in terms of difference in propagation pathways, speed of propagation, and age at propagation onset. Longitudinal MRI datasets of patients with mild cognitive impairments who converted to AD are used to reconstruct the cortical atrophy propagation across all disease stages. Each observation is considered as a signal spatially distributed on a network, such as the cortical mesh, each cortex location being associated to a node. We consider how the temporal profile of the signal varies across the network nodes. We introduce a statistical mixed-effect model to describe the evolution of the cortex alterations. To ensure a spatiotemporal smooth propagation of the alterations, we introduce a constrain on the propagation signal in the model such that neighboring nodes have similar profiles of the signal changes. Our generative model enables the reconstruction of personalized patterns of the neurodegenerative spread, providing a way to estimate disease progression stages and predict the age at which the disease will be diagnosed. The model shows that, for instance, APOE carriers have a significantly higher pace of cortical atrophy but not earlier atrophy onset.

More details in [19].

## 7.7. A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds

**Participants:** Maxime Louis, Benjamin Charlier, Paul Jusselin, Susovan Pal, Stanley Durrleman.

Parallel transport on Riemannian manifolds allows one to connect tangent spaces at different points in an isometric way and is therefore of importance in many contexts, such as for statistics on manifolds. The existing methods to compute parallel transport require either the computation of Riemannian logarithms, such as the Schild's ladder, or the Christoffel symbols. The Logarithm is rarely given in closed form, and therefore costly to compute whereas the number of Christoffel symbols explodes with the dimension of the manifold, making both these methods intractable. From an identity between parallel transport and Jacobi fields, we propose a numerical scheme to approximate the parallel transport along a geodesic. We find and prove an optimal convergence rate for the scheme, which is equivalent to Schild's ladder's. We investigate potential variations of the scheme and give experimental results on the Euclidean two-sphere and on the manifold of symmetric positive-definite matrices.

More details in [23].

## 7.8. Reduction of recruitment costs in preclinical AD trials. Validation of automatic pre-screening algorithm for brain amyloidosis.

**Participants:** Manon Ansart, Stéphane Epelbaum, Geoffroy Gagliardi, Olivier Colliot, Didier Dormont, Bruno Dubois, Harald Hampel, Stanley Durrleman [Correspondant].

We propose a method for recruiting asymptomatic Amyloid positive individuals in clinical trials, using a two-step process. We first select during a pre-screening phase a subset of individuals which are more likely to be amyloid positive based on the automatic analysis of data acquired during routine clinical practice, before doing a confirmatory PET-scan to these selected individuals only. This method leads to an increased number of recruitments and to a reduced number of PET-scans, resulting in a decrease in overall recruitment costs. We validate our method on 3 different cohorts, and consider 5 different classification algorithms for the pre-screening phase. We show that the best results are obtained using solely cognitive, genetic and socio-demographic features, as the slight increased performance when using MRI or longitudinal data is balanced by the cost increase they induce. We show that the proposed method generalizes well when tested on an independent cohort, and that the characteristics of the selected set of individuals are identical to the characteristics of a population selected in a standard way. The proposed approach shows how Machine Learning can be used effectively in practice to optimize recruitment costs in clinical trials.

More details in [3].

## 7.9. Multiplex core-periphery organization of the human connectome

**Participants:** Federico Battiston, Jeremy Guillon, Mario Chavez, Vito Latora, Fabrizio de Vico Fallani [Correspondant].

What is the core of the human brain is a fundamental question that has been mainly addressed by studying the anatomical connections between differently specialized areas, thus neglecting the possible contributions from their functional interactions. While many methods are available to identify the core of a network when connections between nodes are all of the same type, a principled approach to define the core when multiple types of connectivity are allowed is still lacking. Here, we introduce a general framework to define and extract the core–periphery structure of multi-layer networks by explicitly taking into account the connectivity patterns at each layer. We first validate our algorithm on synthetic networks of different size and density, and with tunable overlap between the cores at different layers. We then use our method to merge information from structural and functional brain networks, obtaining in this way an integrated description of the core of the human connectome. Results confirm the role of the main known cortical and subcortical hubs, but also suggest the presence of new areas in the sensori-motor cortex that are crucial for intrinsic brain functioning. Taken together these findings provide fresh evidence on a fundamental question in modern neuroscience and offer new opportunities to explore the mesoscale properties of multimodal brain networks.

More details in [6].

## 7.10. Integrating EEG and MEG signals to improve motor imagery classification in brain-computer interfaces

**Participants:** Marie-Constance Corsi, Mario Chavez, Denis Schwartz, Laurent Hugueville, Ankit Khambhati, Danielle Bassett, Fabrizio de Vico Fallani [Correspondant].

We adopted a fusion approach that combines features from simultaneously recorded electroencephalogram (EEG) and magnetoencephalogram (MEG) signals to improve classification performances in motor imagery-based brain–computer interfaces (BCIs). We applied our approach to a group of 15 healthy subjects and found a significant classification performance enhancement as compared to standard single-modality approaches in the alpha and beta bands. Taken together, our findings demonstrate the advantage of considering multimodal approaches as complementary tools for improving the impact of noninvasive BCIs.

More details in [10].

## 7.11. Role of inter-hemispheric connections in functional brain networks

**Participants:** Johann Martinez [Correspondant], Javier Buldu, David Papo, Fabrizio de Vico Fallani, Mario Chavez.

Today the human brain can be modeled as a graph where nodes represent different regions and links stand for statistical interactions between their activities as recorded by different neuroimaging techniques. Empirical studies have lead to the hypothesis that brain functions rely on the coordination of a scattered mosaic of functionally specialized brain regions (modules or sub-networks), forming a web-like structure of coordinated assemblies (a network of networks). The study of brain dynamics would therefore benefit from an inspection of how functional sub-networks interact between them. In this paper, we model the brain as an interconnected system composed of two specific sub-networks, the left (L) and right (R) hemispheres, which compete with each other for centrality, a topological measure of importance in a networked system. Specifically, we considered functional brain networks derived from high-density electroencephalographic (EEG) recordings and investigated how node centrality is shaped by interhemispheric connections. Our results show that the distribution of centrality strongly depends on the number of functional connections between hemispheres and the way these connections are distributed. Additionally, we investigated the consequences of node failure on hemispherical centrality, and showed how the abundance of inter-hemispheric links favors the functional balance of centrality distribution between the hemispheres.

More details in [25].

## 7.12. Statistical shape analysis of large datasets based on diffeomorphic iterative centroids

**Participants:** Claire Cury, Joan Glaunès, Olivier Colliot.

We proposed an approach for template-based shape analysis of large datasets, using diffeomorphic centroids as atlas shapes. Diffeomorphic centroid methods fit in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework and use kernel metrics on currents to quantify surface dissimilarities. The statistical analysis is based on a Kernel Principal Component Analysis (Kernel PCA) performed on the set of initial momentum vectors which parametrize the deformations. We tested the approach on different datasets of hippocampal shapes extracted from brain magnetic resonance imaging (MRI), compared three different centroid methods and a variational template estimation. The largest dataset is composed of 1,000 surfaces, and we are able to analyse this dataset in 26 h using a diffeomorphic centroid. Our experiments demonstrate that computing diffeomorphic centroids in place of standard variational templates leads to similar shape analysis results and saves around 70% of computation time. Furthermore, the approach is able to adequately capture the variability of hippocampal shapes with a reasonable number of dimensions, and to predict anatomical features of the hippocampus, only present in 17% of the population, in healthy subjects.

More details in [12].

## 7.13. Multi-modal brain fingerprinting: a manifold approximation based framework

**Participants:** Kuldeep Kumar, Olivier Colliot, Christian Desrosiers.

We proposed an efficient framework, based on manifold approximation, for generating brain fingerprints from multi-modal data. The proposed framework represents images as bags of local features, which are used to build a subject proximity graph. Compact fingerprints are obtained by projecting this graph in a low-dimensional manifold, using spectral embedding. Experiments using the T1/T2-weighted MRI, diffusion MRI, and resting state fMRI data of 945 Human Connectome Project subjects demonstrate the benefit of combining multiple modalities, with multi-modal fingerprints more discriminative than those generated from individual modalities. Results also highlight the link between fingerprint similarity and genetic proximity, monozygotic twins having more similar fingerprints than dizygotic or non-twin siblings. This link is also reflected in the differences of feature correspondences between twin/sibling pairs, occurring in major brain structures and across hemispheres. The robustness of the proposed framework to factors like image alignment and scan resolution, as well as the reproducibility of results on retest scans, suggest the potential of multi-modal brain fingerprinting for characterizing individuals in a large cohort analysis. In addition, taking inspiration from

the computer vision community, the proposed rank retrieval evaluation based on the task of twin/sibling identification and using Mean Average Precision (MAP) can be used for a standardized comparison of future brain fingerprints.

More details in [20].

## 7.14. Structural, Microstructural, and Metabolic Alterations in Primary Progressive Aphasia Variants

**Participants:** Alexandre Routier [Correspondant], Marie-Odile Habert, Olivier Colliot, Marc Teichmann.

Neuroimaging studies have described the brain alterations in primary progressive aphasia (PPA) variants (semantic, logopenic, nonfluent/agrammatic). However, few studies combined T1, FDG-PET, and diffusion MRI techniques to study atrophy, hypometabolism, and tract alterations across the three PPA main variants. We therefore explored a large early-stage cohort of semantic, logopenic and nonfluent/agrammatic variants (N = 86) and of 23 matched healthy controls with anatomical MRI (cortical thickness), FDG PET (metabolism) and diffusion MRI (white matter tracts analyses), aiming at identifying cortical and sub-cortical brain alterations, and confronting these alterations across imaging modalities and aphasia variants. In the semantic variant, there was cortical thinning and hypometabolism in anterior temporal cortices, with left-hemisphere predominance, extending toward posterior temporal regions, and affecting tracts projecting to the anterior temporal lobes (inferior longitudinal fasciculus, uncinate fasciculus) and tracts projecting to or running nearby posterior temporal cortices: (superior longitudinal fasciculus, inferior frontal-occipital fasciculus). In the logopenic variant metabolic alterations were more extensive than atrophy affecting mainly the left temporal-parietal junction and extending toward more anterior temporal cortices. Metabolic and tract data were coherent given the alterations of the left superior and inferior longitudinal fasciculus and the left inferior frontal-occipital fasciculus. In the nonfluent/agrammatic variant cortical thinning and hypometabolism were located in the left frontal cortex but Broca's area was only affected on metabolic measures. Metabolic and tract alterations were coherent as reflected by damage to the left uncinate fasciculus connecting with Broca's area. Our findings provide a full-blown statistically robust picture of brain alterations in early-stage variants of primary progressive aphasia which has implications for diagnosis, classification and future therapeutic strategies. They demonstrate that in logopenic and semantic variants patterns of brain damage display a non-negligible overlap in temporal regions whereas they are substantially distinct in the nonfluent/agrammatic variant (frontal regions). These results also indicate that frontal networks (combinatorial syntax/phonology) and temporal networks (lexical/semantic representations) constitute distinct anatomo-functional entities with differential vulnerability to degenerative processes in aphasia variants. Finally, the identification of the specific damage patterns could open an avenue for trans-cranial stimulation approaches by indicating the appropriate target-entry into the damaged language system.

More details in [29].

## 7.15. Neurite density is reduced in the presymptomatic phase of C9orf72 disease

**Participants:** Junhao Wen, Hui Zhang, Daniel Alexander, Stanley Durrleman, Olivier Colliot, Isabelle Le Ber, Anne Bertrand [Correspondant].

In this study, we aimed to assess the added value of neurite orientation dispersion and density imaging (NODDI) compared to conventional DTI and anatomical MRI to detect changes in presymptomatic carriers of chromosome 9 open reading frame 72 (C9orf72) mutation. The PREV-DEMALS study is a prospective, multicenter, observational study of first-degree relatives of individuals carrying the C9orf72 mutation. Sixty-seven participants (38 presymptomatic C9orf72 mutation carriers [C9+], 29 non carriers [C9-]) were included in the present cross-sectional study. Each participant underwent one single-shell, multi-shell diffusion MRI and 3DT1 MRI. Volumetric measures, DTI and NODDI metrics were calculated within regions of interest. Differences in white matter integrity, gray matter volume and free water fraction between C9+ and C9-

individuals were assessed using linear mixed-effects models. Compared with C9-, C9+ demonstrated white matter abnormalities in 10 tracts with neurite density index, and only 5 tracts with DTI metrics. Effect size was significantly higher for the neurite density index than for DTI metrics in two tracts. No tract had a significantly higher effect size for DTI than for NODDI. For gray matter cortical analysis, free water fraction was increased in 13 regions in C9+, whereas 11 regions displayed volumetric atrophy. In conclusion, NODDI provides higher sensitivity and greater tissue-specificity compared to conventional DTI for identifying white matter abnormalities in the presymptomatic C9orf72 carriers. Our results encourage the use of neurite density as biomarker of the preclinical phase.

More details in [34].

## 7.16. Learning myelin content in multiple sclerosis from multimodal MRI through adversarial training

**Participants:** Wen Wei, Emilie Poirion, Benedetta Bodini, Stanley Durrleman, Nicholas Ayache, Bruno Stankoff, Olivier Colliot [Correspondant].

Multiple sclerosis (MS) is a demyelinating disease of the central nervous system (CNS). A reliable measure of the tissue myelin content is therefore essential to understand the physiopathology of MS, track progression and assess treatment efficacy. Positron emission tomography (PET) with [ 11 C]PIB has been proposed as a promising biomarker for measuring myelin content changes in-vivo in MS. However, PET imaging is expensive and invasive due to the injection of a radioactive tracer. On the contrary, magnetic resonance imaging (MRI) is a non-invasive, widely available technique, but existing MRI sequences do not provide, to date, a reliable, specific, or direct marker of either demyelination or remyelination. In this work, we therefore propose Sketcher-Refiner Generative Adversarial Networks (GANs) with specifically designed adversarial loss functions to predict the PET-derived myelin content map from a combination of MRI modalities. The prediction problem is solved by a sketch-refinement process in which the sketcher generates the preliminary anatomical and physiological information and the refiner refines and generates images reflecting the tissue myelin content in the human brain. We evaluated the ability of our method to predict myelin content at both global and voxel-wise levels. The evaluation results show that the demyelination in lesion regions and myelin content in normal-appearing white matter (NAWM) can be well predicted by our method. The method has the potential to become a useful tool for clinical management of patients with MS.

More details in [40].

## 7.17. COGEVIS: A New Scale to Evaluate Cognition in Patients with Visual Deficiency

**Participants:** Claire Meyniel, Dalila Samri, Farah Stefano, Joel Crevoisier, Florence Bonté, Raffaela Migliaccio, Laure Delaby, Anne Bertrand, Marie-Odile Habert, Bruno Dubois, Baram Bodaghi, Stéphane Epelbaum [Correspondant].

We evaluated the cognitive status of visually impaired patients referred to low vision rehabilitation (LVR) based on a standard cognitive battery and a new evaluation tool, named the COGEVIS, which can be used to assess patients with severe visual deficits. We studied patients aged 60 and above, referred to the LVR Hospital in Paris. Neurological and cognitive evaluations were performed in an expert memory center. Thirty-eight individuals, 17 women and 21 men with a mean age of 70.3(SD=1.3 years) and a mean visual acuity of 0.12(SD=0.02), were recruited over a one-year period. Sixty-three percent of participants had normal cognitive status. Cognitive impairment was diagnosed in 37.5% of participants. The COGEVIS score cutoff point to screen for cognitive impairment was 24 (maximum score of 30) with a sensitivity of 66.7% and a specificity of 95%. Evaluation following 4 months of visual rehabilitation showed an improvement of Instrumental Activities of Daily Living ($p = 0.004$), National Eye Institute Visual Functioning Questionnaire ($p = 0.035$), and Montgomery-Åsberg Depression Rating Scale ($p = 0.037$). This study introduces a new short test to screen for cognitive impairment in visually impaired patients.

More details in [27].

## 7.18. Neural correlates of episodic memory in the Memento cohort

**Participants:** Stéphane Epelbaum [Correspondant], Vincent Bouteloup, Jean François Mangin, Valentina La Corte, Raffaela Migliaccio, Hugo Bertin, Marie Odile Habert, Clara Fischer, Chabha Azouani, Ludovic Fillon, Marie Chupin, Bruno Vellas, Florence Pasquier, Frederic Blanc, Audrey Gabelle, Mathieu Ceccaldi, Pierre Krolak-Salmon, Jacques Hugon, Olivier Hanon, Olivier Rouaud, Renaud David, Genevieve Chene, Bruno Dubois, Carole Dufouil.

The free and cued selective reminding test is used to identify memory deficits in mild cognitive impairment and demented patients. It allows assessing three processes: encoding, storage, and recollection of verbal episodic memory. We investigated the neural correlates of these three memory processes in a large cohort study. The Memento cohort enrolled 2323 outpatients presenting either with subjective cognitive decline or mild cognitive impairment who underwent cognitive, structural MRI and, for a subset, fluorodeoxyglucose-positron emission tomography evaluations. Encoding was associated with a network including parietal and temporal cortices; storage was mainly associated with entorhinal and parahippocampal regions, bilaterally; retrieval was associated with a widespread network encompassing frontal regions. The neural correlates of episodic memory processes can be assessed in large and standardized cohorts of patients at risk for Alzheimer's disease. Their relation to pathophysiological markers of Alzheimer's disease remains to be studied.

## 7.19. Cognitive and neuroimaging features and brain amyloidosis in individuals at risk of Alzheimer's disease

**Participants:** Bruno Dubois [Correspondant], Stéphane Epelbaum, Francis Nyasse, Hovagim Bakardjian, Geoffroy Gagliardi, Olga Uspenskaya, Marion Houot, Simone Lista, Federica Cacciamani, Marie Claude Potier, Anne Bertrand, Foudil Lamari, Habib Benali, Jean François Mangin, Olivier Colliot, Remy Genthon, Marie-Odile Habert, Harald Hampel.

Improved understanding is needed of risk factors and markers of disease progression in preclinical Alzheimer's disease. We assessed associations between brain amyloidosis and various cognitive and neuroimaging parameters with progression of cognitive decline in individuals with preclinical Alzheimer's disease. The INSIGHT-preAD is an ongoing single-centre observational study at the Salpêtrière Hospital, Paris, France. Eligible participants were age 70-85 years with subjective memory complaints but unimpaired cognition and memory (Mini-Mental State Examination [MMSE] score $\geq$ 27, Clinical Dementia Rating score 0, and Free and Cued Selective Reminding Test [FCSRT] total recall score $\geq$ 41). We stratified participants by brain amyloid deposition on 18F-florbetapir PET (positive or negative) at baseline. All patients underwent baseline assessments of demographic, cognitive, and psychobehavioural, characteristics, APOE $\epsilon 4$ allele carrier status, brain structure and function on MRI, brain glucose-metabolism on 18F-fluorodeoxyglucose (18F-FDG) PET, and event-related potentials on electroencephalograms (EEGs). Actigraphy and CSF investigations were optional. Participants were followed up with clinical, cognitive, and psychobehavioural assessments every 6 months, neuropsychological assessments, EEG, and actigraphy every 12 months, and MRI, and 18F-FDG and 18F-florbetapir PET every 24 months. We assessed associations of amyloid deposition status with test outcomes at baseline and 24 months, and with clinical status at 30 months. Progression to prodromal Alzheimer's disease was defined as an amnestic syndrome of the hippocampal type. From May 25, 2013, to Jan 20, 2015, we enrolled 318 participants with a mean age of 76·0 years (SD 3·5). The mean baseline MMSE score was 28·67 (SD 0·96), and the mean level of education was high (score >6 [SD 2] on a scale of 1-8, where 1=infant school and 8=higher education). 88 (28showed amyloid deposition and the remainder did not. The amyloid subgroups did not differ for any psychobehavioural, cognitive, actigraphy, and structural and functional neuroimaging results after adjustment for age, sex, and level of education More participants positive for amyloid deposition had the APOE $\epsilon 4$ allele (33 [38%] vs 29 [13%], p<0·0001). Amyloid concentration in CSF significantly correlated with mean 18F-florbetapir uptake at baseline (r=-0·62, p<0·0001) and the ratio of amyloid to amyloid (r=-0·61, p<0·0001), and identified amyloid deposition status with high accuracy (mean area under

the curve values 0·89, 95% CI 0·80-0·98 and 0·84, 0·72-0·96, respectively). No difference was seen in MMSE (28·3 [SD 2·0] vs 28·9 [1·2], p=0·16) and Clinical Dementia Rating scores (0·06 [0·2] vs 0·05 [0·3]; p=0·79) at 30 months (n=274) between participants positive or negative for amyloid. Four participants (all positive for amyloid deposition at baseline) progressed to prodromal Alzheimer's disease. They were older than other participants positive for amyloid deposition at baseline (mean 80·2 years [SD 4·1] vs 76·8 years [SD 3·4]) and had greater 18F-florbetapir uptake at baseline (mean standard uptake value ratio 1·46 [SD 0·16] vs 1·02 [SD 0·20]), and more were carriers of the APOE $\epsilon 4$ allele (three [75%] of four vs 33 [39%] of 83). They also had mild executive dysfunction at baseline (mean FCSRT free recall score 21·25 [SD 2·75] vs 29·08 [5·44] and Frontal Assessment Battery total score 13·25 [1·50] vs 16·05 [1·68]). Brain amyloidosis alone did not predict progression to prodromal Alzheimer's disease within 30 months. Longer follow-up is needed to establish whether this finding remains consistent.

More details in [13].

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Grants with Industry

### *8.1.1. Carthera*

**Participants:** Stéphane Epelbaum [Correspondant], Alexandre Carpentier, Anne Bertrand, Marie Odile Habert.

Project title: Open label phase 1/2 study evaluating the safety and usefulness of transient opening of the blood-brain barrier using low intensity pulsed ultrasounds generated by the implantable device SONOCLOUD in patients with mild Alzheimer's disease

Started in 2016

Amount: 400 K€

Coordinator: Stéphane Epelbaum

Other partners: UPMC, AP-HP

Abstract: This project aims at opening the blood brain barrier (BBB) in 10 mild Alzheimer's disease patients in order to improve the clearance of beta-amyloid and tau deposits in their brain as suggested in mice models of the disease. This first in man study will evaluate the safety and efficacy of an implanted device, SONOCLOUD, to open the BBB 7 times in each participant. Efficacy will be evaluated on the ability of the method to decrease the amyloid load evidenced by AV45 Positron Emission Tomography (PET), increase the brain metabolism analyzed by Fluorodeoxyglucose PET and improve cognition. If successful, this study will pave the way for future trials in which drugs can be used in addition to BBB opening to maximize their effect.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR*

#### *9.1.1.1. ANR-NIH-NSF NETBCI*

**Participants:** Fabrizio de Vico Fallani [Correspondant], Mario Chavez, Denis Schwartz.

Project acronym: NETBCI

Project title: Modeling and predicting brain-computer interface learning from dynamic networks

Duration: Avr 2016 - Avr 2020

Amount: 322k€

Coordinator: Fabrizio De Vico Fallani

Other partners: Complex system group, UPenn, USA

Abstract: This project will bring together expertise in computational and experimental neuroscience, signal processing and network science, statistics, modeling and simulation, to establish innovative methods to model and analyze temporally dynamic brain networks, and to apply these tools to develop predictive models of brain-computer interface (BCI) skill acquisition that can be used to improve performance. Leveraging experimental data and interdisciplinary theoretical techniques, this project will characterize brain networks at multiple temporal and spatial scales, and will develop models to predict the ability to control the BCI as well as methods to engineer BCI frameworks for adapting to neural plasticity. This project will enable a comprehensive understanding of the neural mechanisms of BCI learning, and will foster the design of viable BCI frameworks that improve usability and performance.

*9.1.1.2. ANR-NIH-NSF HIPLAY7*

**Participants:** Olivier Colliot [Correspondant], Marie Chupin, Stanley Durrleman, Anne Bertrand.

Project acronym: HIPLAY7

Project title: Hippocampal layers: advanced computational anatomy using very high resolution MRI at 7 Tesla in humans

Duration: Jan 2017 - Jan 2020

Amount: 770k€

Coordinator: Olivier Colliot and Pierre-François Van de Moortele

Other partners: University of Minnesota, Neurospin

Abstract: The overall goal of this proposal is to develop a coherent mathematical framework for computational anatomy of the internal structures of the hippocampus based on cutting edge MRI acquisition techniques at 7 Tesla. These mathematical and computational approaches are expected to significantly advance the field of computational anatomy of the human brain, breaking down the millimeter barrier of conventional brain morphometry and providing a coherent analysis framework for anatomical data at ultra-high spatial resolution.

*9.1.1.3. ANR PREV-DEMALS*

**Participants:** Olivier Colliot [Correspondant], Marie Chupin, Stanley Durrleman, Anne Bertrand.

Project acronym: PREV-DEMALS

Project title: Predict to prevent frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS)

Duration: Avr 2015 - Avr 2019

Amount: 487k€

Coordinator: Isabelle Le Ber

Other partners: ICM, AP-HP, CHR de Lille, CHU Limoges, CHU Rouen, Laboratory of Biomedical Imaging

Abstract: The project focuses on C9ORF72, the most frequent genetic form of frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS). Since 2006, major discoveries have helped elucidate the pathological bases and linked FTLD and ALS: 1) TDP-43 aggregates in neurons and 2) C9ORF72 mutations in both disorders. Two major pathological subtypes are now defined in FTLD, FTLD-TDP and FTLD-TAU. C9ORF72 mutations (associated to FTLD-TDP) are the most frequent genetic causes of FTLD (15%), FTLD-ALS (65%) and ALS (40%). No curative treatment actually exists, but therapeutics emerged against tau aggregation. The objectives of the project are to develop appropriate cognitive, brain imaging markers and peripheral biomarkers of the early phase of FTLD, to follow disease progression and to guide future targeted therapeutic trials. To address this

questions, we will conduct a multimodal study (cognition, brain structural MRI, brain metabolism - FDG-PET) in C9ORF72 families. The cohort will be followed at 3-time points (M0, M18, M36). Longitudinal analyses will aim at characterizing the trajectory of decline across time. Brain structural changes will be evaluated by 1) morphometric analysis to assess global brain atrophy, cortical thickness and study of the cortical sulci; 2) functional connectivity analysis of resting-state MR data; 3) structural connectivity analysis of diffusion-weighted MRI. Brain metabolism will be evaluated with FDG-PET. We will use the most recent RNA sequencing technology to detect gene expression and RNA splicing alterations in lymphocytes of patients and presymptomatic carriers. The discovery of new markers involved in FTLD will have practical consequences for early and accurate diagnosis of FLD and ALS disease.

*9.1.1.4. ANR IVMRS*

**Participants:** Anne Bertrand [Correspondant], Alexandra Petiet, Mathieu Santin, Francesca Branzoli, Benoit Delatour, Marc Sanson.

Project acronym: IVMRS

Project title: Implantable miniaturized probe for In-vivo Magnetic Resonance Spectroscopy: Application to Murine models of Alzheimer's disease and Gliomas.

Duration: Oct 2016 - Oct 2020

Amount: 633k€

Coordinator: Luc Hebrard

Other partners: ICube - Unistra, Strasbourg; ISA Laboratory, Lyon; NYU School of Medicine, NY, USA.

Abstract: During the development of new therapeutics against brain diseases, the pre-clinical phase, i.e. the validation of treatment delivery, safety and efficacy in animal models of the disease, represents a crucial step. Magnetic Resonance Imaging (MRI) is a method of particular interest at this stage, as it provides non-invasive surrogate endpoints that can help selecting appropriate candidates during the process of drug development. Single Voxel Magnetic Resonance Spectroscopy (SVS) provides non-invasive, in-vivo quantitative measurements of brain metabolites, which reflects functional changes at the cellular and subcellular levels, and can be repeated longitudinally. As high-field MRI has become the benchmark in preclinical research on animal models, it appears possible to investigate the cerebral metabolomics changes in animals, and to use it as a surrogate marker in preclinical therapeutic trials. However, the number of relevant metabolites is much higher than the low number of measurable metabolites with conventional in-vivo high-field SVS. Moreover, considering also the subtle changes of these metabolites at the early stage of the disease, the use of conventional high-field SVS in preclinical studies remains strongly limited. The high volume of the Voxel-of-Interest (VOI), ranging from 10 to 30mm3, which is required to have a usable signal in conventional SVS, and the inherent variability of longitudinal SVS measurement due to the variable position of the VOI in the successive experiments, remain the two major issues when looking during time for small changes in metabolic concentrations and metabolites ratios in a specific small region of the animal brain. The IvMRS project aims at filling this gap by developing the first chronic implantable MRS micro-probe, minimally invasive, exhibiting very high signal sensitivity, and sharp spectral peaks, from sub-millimetric VOI. Such a probe will allow detecting a much higher number of metabolites than conventional in-vivo SVS. The probe will work at frequencies ranging from 300MHz to 500MHz in ultra-high field Magnetic Resonance Imaging scanners, 7T and 11.7T. It will embed a specific micro-coil antenna, a low-noise signal conditioning circuit designed in CMOS microelectronics technology, as well as an accurate on-chip positioning sensor. It will be dedicated to the study of changes in brain metabolite markers of two major diseases, Alzheimer's disease and cerebral gliomas, and to the assessment of effective therapeutic strategies.

### *9.1.2. Inria Project Labs*

*9.1.2.1. IPL Neuromarkers*

**Participants:** Stanley Durrleman [Correspondant], Olivier Colliot [Correspondant], Fabrizio de Vico Fallani, Anne Bertrand, Stéphane Epelbaum.

Project acronym: Neuromarkers

Project title: Design of imaging biomarkers of neurodegenerative diseases for clinical trials and study of their genetic associations

Duration: 2017-2021

Coordinators: Stanley Durrleman and Olivier Colliot

Other partners: Inria GENSCALE, Inria BONSAI, Inria DYLISS, Inria XPOP, ICM, IHU/ICM iConics

Abstract: The Inria Project Lab Neuromarkers aims to develop new statistical and computational approaches to integrate multimodal imaging and omics data and to demonstrate their potential to identify early alterations and predict progression of neurodegenerative diseases. To tackle this challenge, the project brings together multidisciplinary expertise from Inria and ICM (Brain and Spine Institute) in the fields of statistical learning, brain imaging, bioinformatics, knowledge modeling, genomics and neurodegenerative diseases.

### *9.1.3. IHU*

*9.1.3.1. General program*

**Participants:** Olivier Colliot, Stanley Durrleman, Didier Dormont, Ninon Burgos, Stéphane Epelbaum, Fabrizio de Vico Fallani.

Project acronym: IHU-A-ICM

Project title: Institute of Translational Neuroscience

Founded in 2011

General Director: Bertrand Fontaine

The IHU-A-ICM program was selected, in 2011, in a highly competitive national call for projects. A 10-year, 55M€ program, has been implemented by a recently created foundation for scientific cooperation. Based on the clinical and scientific strenghts of the ICM and the hospital Department of Nervous System Diseases, it mainly supports neuroscience research, but is also invested in improving care and teaching. ARAMIS is strongly involved in the IHU-A-ICM project, in particular in WP6 (neuroimaging and electrophysiology), WP7 (biostatistics), WP2 (Alzheimer) and WP5 (epilepsy). We have started collaborations with the new bioinformatics/biostatistics platform (IHU WP7, head: Ivan Moszer), in particular through a joint project on the integration of imaging and genomics data.

*9.1.3.2. ICM-Internal Research projects*

**Participants:** Anne Bertrand [Correspondant], Takoua Kaaouana, Benoit Delatour, Alexandra Petiet, Olivier Colliot, Arnaud Marcoux.

Project title: The Histo-MRI project: targeting MR signature of tauopathy from micro- to macroscopy

Started in 2014

Coordinator: Anne Bertrand

Identifying morphological MR signatures of brain diseases usually follows a top-down process, which starts by describing a pattern of MR signal changes in patients, hypothesizes an underlying pathological mechanism, and confirms this mechanism by correlating the observed MR signal changes with histological lesions on post-mortem examination. This top-down process, relevant for large, centimetric brain lesions, becomes inappropriate when targeting the MR signal intensity changes associated with microscopic lesions. Our project aims at developing an MR biomarker of NFT using a new bottom-up approach. We will start by identifying the MR signal changes associated

with the presence of NFT at the level of the histological slice, and utilize these findings to develop a method of NFT quantification on clinical, millimetric 3D MR images. To achieve this goal, we will develop and implement a 11.7T histological coil dedicated to the scanning of histological slices, which allows both ultra-high resolution MR imaging (up to 33 microns in-plane) and perfect co-registration with histological staining, performed subsequently on the same slice. This method has the potential to provide a novel biomarker of tauopathy that could not have been identified using the usual top-down approach. It also envisions the possibility to describe and understand new MRI contrasts in other neurodegenerative diseases associated with microscopic deposition of various proteins.

*9.1.3.3. ICM-Internal Research projects*

**Participants:** Mario Chavez, Fabrizio de Vico Fallani [Correspondant].

Project title: Non-invasive manipulation of brain synchrony to enhance brain function and rehabilitate faulty cognition in humans: A proof of concept

Started in 2014

Coordinator: Antoni Valero Cabre (ICM-team "Dynamiques Cérébrales, Plasticité et Rééducation")

Other partners: Service des Urgences Cérébro-Vasculaires de l'Hôpital Pitié-Salpêtrière, Paris.

The long-term goal of this project is to develop the use of non-invasive manipulation of abnormal cerebral oscillations underlying cognitive activity to restore brain function in neurological patients. Cognitive functions emerge from large distributed networks organized in space and time. The short-term goal of this application is to study the causal role played by oscillatory activity in visual awareness and test whether their manipulation by non-invasive brain stimulation has the potential to restore its function in stroke patients.

*9.1.3.4. ICM BBT Program - project PredictICD*

**Participants:** Olivier Colliot [Correspondant], Jean-Christophe Corvol [Correspondant], Johann Faouzi.

Project title: Predict impulse control disorders in Parkinson's disease (PREDICT-ICD)

Started in 2018

Coordinators: Olivier Colliot and Jean-Christophe Corvol (ICM)

In Parkinson's disease (PD), the therapeutic strategy is based on the dopamine replacement therapy. Although available since the 1960s', it is only relatively recently that behavioral disorders associated with these drugs have been described. Gathered under the term of "behavioral addiction", they include impulse control disorders (ICDs), dopamine dysregulation syndrome (DDS), and punding. Interestingly, whereas addiction to L-dopa itself occurs quasi exclusively with L-dopa, ICDs appear electively under dopamine agonist (DA) therapy. The objectives of this project are: i) to elucidate the genetic basis of DA induced ICDs in PD patients from several international cohorts; ii) to develop and validate a machine learning model to predict the occurrence of ICDs from the combination of clinical and genetic data.

*9.1.3.5. ICM BBT Program - project DYNAMO*

**Participants:** Stanley Durrleman [Correspondant], Harald Hampel [Correspondant], Sabrina Fontanella, Simone Lista, Olivier Colliot, Stephanie Allassonniere, Jean-Baptiste Schiratti, Bruno Dubois, Hovagim Bakardjian, Remi Genthon, Enrica Cavedo, Katrine Rojkowa.

Project title: Dynamic models of disease progression across Alzheimer's disease stages informed by multimodal neuroimaging and biological data

Started in 2016

Coordinator: Stanley Durrleman and Harald Hampel

Other partners: Institut de la Mémoire et de la maladie d'Alzheimer

The estimation of data-driven models of disease progression for neurodegenerative diseases, including Alzheimer's disease (AD), is crucial to confirm, refine and extend the current hypothetical models. The estimation of such quantitative models from longitudinal data sets is notably difficult because of the lack of principled methodological frameworks for the analysis of spatiotemporal data.

The project builds on an innovative mathematical, statistical, and computational framework to automatically align the dynamics and the direction of individual trajectories of the evolving pathology, and then to infer a normative scenario of disease progression across different disease stages. The estimated scenario will combine spatiotemporal maps of lesion propagation, such as maps of amyloid deposition or cortical atrophy, and global measurements such as levels of CSF biomarkers. It will be possible to estimate not only a normative scenario but also the inter-individual variability in the values, dynamics and direction of both topographical and pathophysiological biomarkers changes during the course of the disease.

The application of this technology to publicly available and in-house longitudinal data sets of individuals from the asymptomatic at risk to the prodromal and dementia stages will yield new insights into the pathophysiology of AD from the preclinical to the AD dementia stages. This quantitative data-driven approach will be exploited to assess and refine the current qualitative hypothetical models of AD progression. Notably, it will complement these models with typical pathways of lesion propagation in the brain during disease progression. It will also highlight the effect of the known risk factors of AD such as apolipoprotein E genotype on the disease progression profile.

The project will open up the concrete possibility to derive a computer-aided diagnosis, staging, and prognosis tool for a better recruitment of patients in clinical studies and to assist clinicians in the diagnosis and the monitoring of both disease progression and treatment efficacy.

*9.1.3.6. ICM BBT Program - project SEMAPHORE*

**Participants:** Stanley Durrleman [Correspondant], Stéphane Lehéricy [Correspondant], Jean-Christophe Corvol, Marie Vidailhet, Raphael Couronné, Safia Said.

Project title: Personalized progression model of Parkinson's disease

Started in 2018

Coordinator: Stanley Durrleman and Stéphane Lehéricy

Other partners: Neurology and Neuro-radiology departments, Pitié-Salpêtrière Hospital, AP-HP

The aim of this project is to build a personalizable model of Parkinson's disease (PD) progression integrating the complex dynamical interplay between phenotypic, imaging, genetic and metabolic alterations. We will identify and validate markers for monitoring of progression of brain damage in early and prodromal PD and identify conversion markers in subjects at risk of PD (idiopathic rapid eye movement sleep behavior disorders iRBD, PD- related mutation carriers). We will describe the appearance, characterize clinical phenotypes of PD, and identify modifier genes of disease phenotype. To this aim, we will relie on a novel statistical learning method using Bayesian non-linear mixed-effects model allowing to combine and realign short term sequence data to estimate a long-term scenario of disease progression. This method is able to estimate individual stages of disease progression and to analyze automatically non-linear spatiotemporal patterns of data change. It estimates both a group-average scenario of PD progression as well as the inter-individual variability of this model in terms of age at onset, pace of disease progression and variability in the spatiotemporal trajectory of data changes. We will analyse the effect of genetic variants in the modulation of these non-linear progression patterns, and assess the statistical power of the individual parameters encoding for these patterns. The method will be applied to two sets of longitudinal data from the local prospective NUCLEIPARK (60 PD patients, 20 patients with iRBD, 60 controls) and ICEBERG studies (200 early idiopathic PD, 50 iRBD, 30 GBA and LRRK2 PD-related mutation carriers, 50 controls). Examinations included clinical, biological, and neurophysiological data, and multimodal 3T MRI, DATScan, and skin and salivary gland biopsies. The models of PD progression

for each category of subjects will be released to the community, as well as the software for reproducibility purposes.

*9.1.3.7. ICM BBT Program - project ATTACK*

**Participants:** Fabrizio de Vico Fallani [Correspondant], Charlotte Rosso [Correspondant], Marie-Constance Corsi, Laurent Hugueville.

Project title: ATTACK Brain Network Models Of Motor Recovery After Stroke

Started in 2018

Coordinator: Fabrizio De Vico Fallani, Charlotte Rosso

Other partners: Neurology and Stroke departments, Pitié-Salpêtrière Hospital, AP-HP

Like in other connected systems, studying the structure of the interactions between different brain regions has profound implications in the comprehension of emergent complex phenomena as, for example, the capability of the human brain to functionally reorganize after cerebrovascular "attacks" or stroke. This dynamic skill, which is known in neuroscience as neural plasticity, is not only interesting from a network science perspective, but it also plays a crucial role in determining the motor/cognitive recovery of patients who survive a stroke. As a critical innovation, this project proposes to develop a systematic and rigorous approach based on neuroimaging techniques, signal processing, and network science for the modeling and analysis of temporally dynamic neural processes that characterize motor recovery after stroke. To achieve these goals, this project is organized around the following objectives: i) acquiring a comprehensive longitudinal dataset of brain and behavioral/clinical data after stroke, ii) developing new analytic tools to characterize and generate temporally dynamic brain networks, iii) building network-based models of motor recovery after stroke, accounting for individual patients. These objectives involve an intensive gathering of heterogeneous mass data, their processing, the subsequent outcome interpretation and statistical simulation, as well as the development of longitudinal models and network-based diagnostics of the patient's motor recovery progress. Results will be first characterized from pure network-theoretic and neuroscience perspectives, so as to highlight fundamental research challenges, and then validated to clarify the importance and the applicability to the clinical scenario. Our results will unveil multiscale properties of dynamic brain networks and identify predictive neuromarkers for motor recovery after stroke. This project has a two-fold impact on the society. On the one hand, it will provide new methods and robust tools to properly characterize and model temporally dynamic networks in neuroscience. On the other hand, it will provide longitudinal models of motor recovery in stroke patients that can potentially unveil the neural substrate that underpins rehabilitation, improve prognosis, and eventually lower cost of hospitalization time. From a broader perspective this interdisciplinary project proposes a transformative approach to analyze large-scale neural systems.

### *9.1.4. National Networks*

- GdR Statistics and Medicine - http://gdr-stat-sante.math.cnrs.fr/spip/
- GdR (MaDICS) Masses de Données, Informations et Connaissances en Sciences Big Data - Data ScienceStatistics and Medicine - http://www.madics.fr/reseaux/
- F. De Vico Fallani participated to the GdR (HANDICAP) in the framework of the future strategy of Inria
- F. De Vico Fallani was founding member of the CORTICO national network for brain-computer interfaces

### *9.1.5. Other National Programs*

*9.1.5.1. Programme Hospitalier de Recherche Clinique (PHRC)*

**Participants:** Olivier Colliot, Stanley Durrleman, Didier Dormont.

- PHRC PredictPGRN, co-funding by Alzheimer Plan, *Caractérisation multimodale prospective de la démence frontotemporale dûe à des mutations du gène PGRN à un stade symptomatique et présymptomatique*. (Coordinator : A. Brice)
- PHRC ImaBio3, co-funding by Roche (pharmaceutical industry), *Rôle des réactions cellulaires sanguines, inflammatoires et immunitaires anti-amyloïde centrales et périphériques dans la maladie d'Alzheimer débutante*. (Coordinator : M. Sarazin)
- PHRC CAPP, *Caractérisation linguistique, anatomique/métabolique et biologique des différentes formes d'aphasie primaire progressive : vers le rationnel pour des essais pharmacologiques et des rééducations du langage ciblées*. (Coordinator: M. Teichmann)

*9.1.5.2. Institut Universitaire d'Ingénierie pour la Santé (IUIS)*

**Participants:** Mario Chavez, Xavier Navarro.

Project acronym: DYSPEV

Project title: Dépistage de la dyspnée par potentiels évoqués visuels

Funded in 2014

Amount: 38K€

Coordinator: Thomas Similowski

Other partners: UPMC, Inserm UMR 1158

Abstract: Steady state visual evoked potentials (SSVEP) have been widely utilized in brain computer interfacing (BCI) in last years. In this project, we explore the possibilities of SSVEP to manage the communication between patients suffering from respiratory disorders and health care providers. By imposing different breathing constraints, we use a SSVEP-based brain computer interface to help those subjects to communicate their breathing sensations (breathing well/breathing bad).

## 9.2. European Initiatives

### *9.2.1. FP7 & H2020 Projects*

*9.2.1.1. H2020 - Project EuroPOND*

**Participants:** Olivier Colliot, Stanley Durrleman, Manon Ansart, Igor Koval, Alexandre Bône.

Project acronym: EuroPOND

Project title: Data-driven models for Progression Of Neurological Disease

Duration: Jan 2016 - Dec 2019

Amount: 6M€

Coordinator: Daniel Alexander

Other partners: University College London (UK), EMC Rotterdam (The Netherlands), VUMC (The Netherlands), Fate Bene Fratelli (Italy), Carol Besta Institute (Italy), Université de Genève (Switzerland), Icometrix (Belgium)

Abstract: EuroPOND will develop a data-driven statistical and computational modeling framework for neurological disease progression. This will enable major advances in differential and personalized diagnosis, prognosis, monitoring, and treatment and care decisions, positioning Europe as world leaders in one of the biggest societal challenges of 21st century healthcare. The inherent complexity of neurological disease, the overlap of symptoms and pathologies, and the high comorbidity rate suggests a systems medicine approach, which matches the specific challenge of this call. We take a uniquely holistic approach that, in the spirit of systems medicine, integrates a variety of clinical and biomedical research data including risk factors, biomarkers, and interactions. Our consortium has a multidisciplinary balance of essential expertise in mathematical/statistical/computational modelling;

clinical, biomedical and epidemiological expertise; and access to a diverse range of datasets for sporadic and well-phenotyped disease types.The project will devise and implement, as open-source software tools, advanced statistical and computational techniques for reconstructing long-term temporal evolution of disease markers from cross-sectional or short-term longitudinal data. We will apply the techniques to generate new and uniquely detailed pictures of a range of important diseases. This will support the development of new evidence-based treatments in Europe through deeper disease understanding, better patient stratification for clinical trials, and improved accuracy of diagnosis and prognosis. For example, Alzheimer's disease alone costs European citizens around €200B every year in care and loss of productivity. No disease modifying treatments are yet available. Clinical trials repeatedly fail because disease heterogeneity prevents bulk response. Our models enable fine stratification into phenotypes enabling more focussed analysis to identify subgroups that respond to putative treatments.

*9.2.1.2. FET Flagship - Human Brain Project*

**Participants:** Olivier Colliot, Stanley Durrleman.

Project acronym: HBP

Project title: Human Brain Project

Sub-project: SP8 - Medical Informatics Platform

Duration: 2016-

Abstract: The Human Brain Project (HBP) is a European Commission Future and Emerging Technologies Flagship. The HBP aims to put in place a cutting-edge, ICT-based scientific Research Infrastructure for brain research, cognitive neuroscience and brain-inspired computing. The Project promotes collaboration across the globe, and is committed to driving forward European industry. Our team is involved in the Subproject SP8 (Medical Informatics Platform). The Medical Informatics Platform (MIP) is an innovative data management system that gives researchers the means to access and analyse large amounts of anonymized clinical neuroscience data. Within that framework, we will develop and implement a method to construct disease progression models from longitudinal biomarkers. The method will use statistical learning techniques to infer a long-term disease progression model from multiple short term data from a series of individuals. The model will account for variability in age at disease onset, pace of disease progression and trajectories of biomarkers changes across individuals in the observed population.

*9.2.1.3. ERC - LEASP*

**Participant:** Stanley Durrleman.

Project acronym: LEASP

Project title: Learning Spatiotemporal Patterns in Longitudinal Image Data Sets of the Aging Brain

Duration: 2016-2021

Abstract: Time-series of multimodal medical images offer a unique opportunity to track anatomical and functional alterations of the brain in aging individuals. A collection of such time series for several individuals forms a longitudinal data set, each data being a rich iconic-geometric representation of the brain anatomy and function. These data are already extraordinary complex and variable across individuals. Taking the temporal component into account further adds difficulty, in that each individual follows a different trajectory of changes, and at a different pace. Furthermore, a disease is here a progressive departure from an otherwise normal scenario of aging, so that one could not think of normal and pathologic brain aging as distinct categories, as in the standard case-control paradigm.

Bio-statisticians lack a suitable methodological framework to exhibit from these data the typical trajectories and dynamics of brain alterations, and the effects of a disease on these trajectories, thus limiting the investigation of essential clinical questions. To change this situation, we propose to construct virtual dynamical models of brain aging by learning typical spatiotemporal patterns of alterations propagation from longitudinal iconic-geometric data sets.

By including concepts of the Riemannian geometry into Bayesian mixed effect models, the project will introduce general principles to average complex individual trajectories of iconic-geometric changes and align the pace at which these trajectories are followed. It will estimate a set of elementary spatiotemporal patterns, which combine to yield a personal aging scenario for each individual. Disease-specific patterns will be detected with an increasing likelihood.

This new generation of statistical and computational tools will unveil clusters of patients sharing similar lesion propagation profiles, paving the way to design more specific treatments, and care patients when treatments have the highest chance of success.

## 9.3. International Initiatives

### *9.3.1. Informal International Partners*

- F. De Vico Fallani has a collaboration with the University Penn, Philadelphia, US (Prof. Danielle Bassett).
- F. De Vico Fallani has a collaboration with the University of Rome, Italy (Prof. Stefania Colonnese).
- O. Colliot has an enduring collaboration with the Center for Magnetic Resonance Research, University of Minnesota, USA (P-F Van de Moortele, T. Henry).
- S. Durrleman and O. Colliot have a collaboration with the Center for Medical Image Computing (CMIC) at University College London (UCL), London, UK (D. Alexander, H. Zhang).

## 9.4. International Research Visitors

### *9.4.1. Visits of International Scientists*

- Dr. Sarah-Christine Villeneuve spent a year from the 4th of December 2017 to the 30th of November 2018 as a clinical research fellow in Pitié Salpêtrière Hospital under the supervision of Stéphane Epelbaum (Sabbatical program).

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

- S. Durrleman served as Program Chair for the eight International Workshop on Biomedical Image Registration (WBIR'18, Leiden, The Netherlands)
- F. De Vico Fallani served as Program Chair for the Network Neuroscience Satellite (NETSCI'18, Paris, France)
- F. De Vico Fallani served as Program Chair for the Brainhack Networks workshop (ICM, Paris, France)

*10.1.1.2. Member of the Organizing Committees*

N. Burgos co-organized the international workshop on Simulation and Synthesis in Medical Imaging (SASHIMI) 2018, a satellite workshop of MICCAI 2018.

S. Epelbaum was session co-chair at the Alzheimer Association International Conference in Chicago 14-18th July 2018

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Member of the Conference Program Committees*

- O. Colliot served as Program Committee member for the international conference SPIE Medical Imaging (Houston, USA, 2018) and for the international workshop PatchMI (Granada, Spain, 2018).
- F. De Vico Fallani served as Program Committee member for the following international conferences: NETSCI (Paris, 2018), Complex Networks (Cambridge, UK, 2018), ComplNet (Zaragoza, Spain, 2018)

*10.1.2.2. Reviewer*

- N. Burgos acted as a reviewer for the international workshops on Deep Learning in Medical Image Analysis (DLMIA) and on Simulation and Synthesis in Medical Imaging (SASHIMI).
- O. Colliot acted as a reviewer for the international conferences SPIE Medical Imaging, Annual meeting of the Organization for Human Brain Mapping (OHBM) and the international workshop PatchMI.

### *10.1.3. Journal*

*10.1.3.1. Member of the Editorial Boards*

- S. Durrleman is associate editor of the journals: IEEE Transactions on Medical Imaging, and Neurons, Behavior, Data analysis, and Theory (NBDT)
- F. De Vico Fallani is associate editor of the journal PLoS One
- O. Colliot is a member of the Editorial Board of the journal Medical Image Analysis (Elsevier).
- S. Epelbaum is member of the editorial board for the "Médecine, Cognition et Vieillissement" Scientific Journal.

*10.1.3.2. Reviewer - Reviewing Activities*

- N. Burgos acted as a reviewer for IEEE Transactions on Medical Imaging; NeuroImage; Medical Image Analysis; Journal of Nuclear Medicine; IEEE Transactions on Radiation and Plasma Medical Sciences; IEEE Journal of Biomedical and Health Informatics; EJNMMI Physics; International Journal of Radiation Oncology, Biology, Physics; Sensors.
- Olivier Colliot acted as a reviewer for Medical Image Analysis, NeuroImage, NeuroImage: Clinical, IEEE Transactions on Medical Imaging.
- F. De Vico Fallani acted as a reviewer for IEEE TNRSE, Neuroimage, PLoS Comp Biol, PLoS One, Cereb Cortex.
- S. Epelbaum acted as a reviewer for Alzheimer's & Dementia, the Journal of Alzheimer's disease, Brain and BMJ Neurology.

### *10.1.4. Invited Talks*

N. Burgos gave an invited lecture at the workshop on Machine Learning in Radiology in Lausanne, Switzerland, November 2018.

N. Burgos gave an invited presentation at the course "Pattern Recognition for Neuroimaging" at the Annual Meeting of the Organization for Human Brain Mapping, Singapore, June 2018.

O. Colliot gave an invited presentation at the Singapore-France Artificial Intelligence Workshop, Singapore, June 2018.

O. Colliot gave an invited presentation at the Netherlands-France "Erasmus-Descartes" Artificial Intelligence Workshop, Paris, France, November 2018.

S. Durrleman gave an keynote presentation at the MICCAI workshop ShapeMI (Shape in medical imaging), Granada, Septembre 2018, and an invited presentation at the Symposium on Multivariate analyses, Modelling and Machine Learning in Neuroimaging Research, University Paul Sabatier of Toulouse.

S. Epelbaum and S. Durrleman gave a plenary lecture at College de France for the colloquium "Imagerie médicale et apprentissage automatique : vers une intelligence artificielle ?"

F. De Vico Fallani gave invited lectures at CuttingEEG International conference (ICM, Paris, 2018)

F. De Vico Fallani gave a plenary lectures at Neurospin CEA Institute (Saclay, France)

### *10.1.5. Scientific Expertise*

- Olivier Colliot is a member of the "Commission des emplois scientifiques" of the Inria Paris Center, in charge of evaluating applications for PhD fellowships, postdoc fellowships and secondments.
- Olivier Colliot acts as an expert for GENCI (the national facility for high-performance computing).

### *10.1.6. Research Administration*

- S. Durrleman serves as the coordinator of the ICM Center for Neuroinformatics, and the scientific director of the iCONICS core facility on data management and analytics.

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master: Olivier Colliot coordinates the course "Méthodes d'imagerie médicale" of the Master 2 in Computer Science of Sorbonne University.

Master: Olivier Colliot, Master in Computer Science, 4.5 hours (eqTD), Sorbonne University

Engineering school: Olivier Colliot, 3 hours (eqTD), Mines ParisTech

Medical school: Didier Dormont is the Director of the University Diploma (DIU) "Diagnostic and Therapeutic Neuroradiology", Université Pierre et Marie Curie

Medical school: Didier Dormont, Courses for Medical Students, Université Pierre et Marie Curie

Medical school: Didier Dormont organizes and participates in the practical teaching of Neuroradiology for Medical Students in the Department of Diagnostic Neuroradiology of Pitié Salpêtrière University Hospital

Medical school: Didier Dormont organizes and participates in the practical teaching of Neuroradiology for Radiology Specializing Residents in the Department of Diagnostic Neuroradiology of Pitié Salpêtrière University Hospital

S. Durrleman, Geometrical approaches in Statistical Learning, 21 hours, Master 2 "Mathématiques, Vision, Apprentissage", ENS Paris Saclay, France.

Master: Stéphane Epelbaum, Master in Neuroscience, 4 hours (eqTD), Université Pierre et Marie Curie

Medical school: Stéphane Epelbaum gives lectures in Neurology on the topic of degenerative diseases for medical students of the UPMC (10 hours/year) and is regional supervisor of the national Inter University Diploma on Alzheimer's disease and Related disorders for Paris since 2015.

### *10.2.2. Supervision*

PhD in progress : Giulia Bassignana, "Identification of driver nodes in biological networks", Inserm, started in 2017, advisors: Fabrizio De Vico Fallani, Olivier Colliot, Violetta Zujovic

PhD in progress : Tiziana Cattai, "Leveraging brain connectivity networks to detect mental states in brain-computer interfaces", Inria, started in 2017, advisor: Fabrizio De Vico Fallani

PhD in progress : Catalina Obando-Forero, "Graph models of cortical plasticity in temporal brain networks", Inria, started in 2015, advisor: Fabrizio De Vico Fallani

PhD in progress : Jeremy Guillon, "Méthode d'analyse multimodale de connectivités neuronales basée sur la théorie des réseaux complexes multicouches", Université Pierre et Marie Curie, started in 2015, advisors: Fabrizio De Vico Fallani and Mario Chavez

PhD Cifre in progress : Fanny Grosselin, "Fouille des données EEG et suivi longitudinal grande échelle pour le diagnostic et la prédiction du niveau de stress chez l'homme", EDITE Université Pierre et Marie Curie, started in 2016, advisors: Fabrizio De Vico Fallani and Mario Chavez,

PhD in progress : Junhao Wen, “Cortical morphometry for discovering new biomarkers of neurodegenerative diseases”, Université Pierre et Marie Curie, Started in 2015, advisors: Olivier Colliot, Anne Bertrand and Stanley Durrleman

PhD in progress : Jorge Samper-Gonzalez, “Learning from heterogeneous data for prediction of Alzheimer’s disease”, Université Pierre et Marie Curie, Started in 2015, advisors: Olivier Colliot and Theodoros Evgeniou

PhD in progress : Alexandre Routier, “Multimodal neuroimaging for characterization of primary progressive aphasias”, Université Pierre et Marie Curie, Started in 2015, advisors: Marc Teichmann, Olivier Colliot and Marie-Odile Habert

PhD in progress: Pascal Lu, “Machine learning from multimodal genetic and neuroimaging data for personalized medicine”, Université Pierre et Marie Curie, Started 2016, advisor: O. Colliot

PhD in progress: Wen Wei, “Learning brain alterations in multiple sclerosis from multimodal neuroimaging data”, Université de Nice Sophia-Antipolis, Started 2016, advisors: N. Ayache, O. Colliot and S. Durrleman

PhD in progress: Alexandre Bône, “Learning methods for the spatiotemporal analysis of longitudinal image data: application to the diagnosis, prognosis and monitoring of Alzheimer’s disease”, started 2016, advisors: O. Colliot and S. Durrleman

PhD in progress: Manon Ansart, “Automatic recommendation systems built on the statistical exploitation of longitudinal medical data sets”, started 2016, advisors: D. Dormont and S. Durrleman

PhD in progress: Maxime Louis, “Learning spatiotemporal trajectories of iconic-geometric data sets”, started 2016, advisors: S. Durrleman

PhD in progress: Igor Koval, “Construction of disease progression models from multimodal longitudinal data”, started 2016, advisors: S. Allassonnière and S. Durrleman

PhD in progress: Raphael Couronné, “Spatiotemporal analysis of the progression of the Parkinson’s Disease informed by multimodal longitudinal data”, advisor: S. Durrleman

PhD in progress: Thomas Lartigue, “Mixture Models in Gaussian Graphical Models”, advisors: S. Allassonnière and S. Durrleman

PhD in progress: Vianney Debavelaere, “Analysis of distribution of spatiotemporal trajectories in heterogeneous populations”, advisors: S. Allassonnière and S. Durrleman

PhD in progress: Lou Albessard, “analyse de la covariation du crâne et de l’endocrâne dans le genre *Homo*”, advisors: D. Grimaud-Hervé and S. Durrleman

PhD in progress: Johann Faouzi, “Machine learning approaches to predict impulse control disorders in Parkinson’s disease”, advisors: O. Colliot and J.-C. Corvol

PhD in progress: Simona Bottani, “Machine learning for differential diagnosis of neurodegenerative diseases from multimodal data”, advisors: O. Colliot and N. Burgos

PhD in progress: Elina Thibeau–Sutre, “Unsupervised learning from neuroimaging data to identify disease subtypes in Alzheimer’s disease and related disorders”, advisors: D. Dormont and N. Burgos

PhD in progress: Federica Cacciamani, “Awareness for cognitive decline in the earliest stages of Alzheimer’s disease”, advisor: S. Epelbaum

### *10.2.3. Juries*

- Olivier Colliot participated, as referee, to the PhD committee of Matthieu Van Houtte (University of Lille).
- Olivier Colliot participated, as examiner, to the HDR committee of Stanley Durrleman (Sorbonne University).
- Olivier Colliot participated to the progress report PhD committee of Ekaterina Kalinicheva (ISEP).

- S. Durrleman participated, as examiner, to the PhD committee of W. Huizinga at Erasmus University Rotterdam, The Netherlands.
- S. Durrleman participated, as co-supervisor, to the PhD committee of L. Albessard at Museum National d'Histoire Naturelle, Paris, France.
- Stéphane Epelbaum participated, as examiner, to the PhD committee of Benoît Souchet (CEA).
- Stéphane Epelbaum participated, as examiner, to the PhD committee of Adrien Julian (University of Poitiers).

## 10.3. Popularization

### *10.3.1. Articles and contents*

- S. Durrleman's interviews were published in the magazines Le Point, Sciences et Avenir, Usine Nouvelle, Le Temps, revue Pharma, and were broadcasted on the radio France Culture. He published also an article in the special issue on artificial intelligence of the newspaper Libération.
- S. Durrleman did popularization presentations at Rotary Club Chamonix, Cercle des amis de l'ICM, and Salon de l'argus de l'assurance.
- F. De Vico Fallani did a popularization presentation at La Française - Investing group, Paris
- Olivier Colliot gave an interview for the TV channel France 5, for the program "Magazine de la Santé.
- Stéphane Epelbaum participated to multiple events dedicated to general audience outreach including: articles in journals (L'Express, Cerveau & Psycho, Sciences Avenir), Radio shows and podcasts (France inter: Grand Bien vous fasse, Figaro Live) and TV shows (Pourquoi docteur, Questions aux experts).
- The team contributed to the book "Le Grand Atlas du Cerveau" (Glénat).

### *10.3.2. Interventions*

- O. Colliot gave a presentation to members of the German Bundestag, Paris, France, December 2018.
- Stéphane Epelbaum participated to a charity event at the Foire internationale d'Art Contemporain (FIAC) 2018 on the 20th of November 2018.

### *10.3.3. Internal action*

- O. Colliot gave a presentation for the charity event ("Evènement solidaire" of Inria, September 2018.

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] J. GUILLON.*Multilayer approach to brain connectivity in Alzheimer's disease*, Pierre and Marie Curie University, November 2018, https://tel.archives-ouvertes.fr/tel-01985286

[2] A. M. ROUTIER.*Multimodal brain imaging for the study of progressive primary aphasia*, Sorbonne Université , UPMC, December 2018, https://tel.archives-ouvertes.fr/tel-01992799

### Articles in International Peer-Reviewed Journal

[3] M. ANSART, S. EPELBAUM, G. GAGLIARDI, O. COLLIOT, D. DORMONT, B. DUBOIS, H. HAMPEL, S. DURRLEMAN.*Reduction of recruitment costs in preclinical AD trials. Validation of automatic pre-screening algorithm for brain amyloidosis*, in "Statistical Methods in Medical Research", January 2019, 096228021882303 [*DOI :* 10.1177/0962280218823036], https://hal.archives-ouvertes.fr/hal-01964942

[4] H. ARABI, J. DOWLING, N. BURGOS, X. HAN, P. GREER, N. KOUTSOUVELIS, H. ZAIDI.*Comparative study of algorithms for synthetic CT generation from MRI: Consequences for MRI-guided radiation planning in the pelvic region*, in "Medical Physics", September 2018 [*DOI :* 10.1002/MP.13187], https://hal.inria.fr/hal-01890646

[5] N. BADAT, C. CHOLET, G. HERVÉ, N. PYATIGORSKAYA, S. TRUNET, D. DORMONT, B. LAW-YE.*Malignant transformation of epidermoid cyst with diffuse leptomeningeal carcinomatosis on skull base and trigeminal perineural spread*, in "Journal de Neuroradiologie / Journal of Neuroradiology", September 2018, vol. 45, n° 5, p. 337-340 [*DOI :* 10.1016/J.NEURAD.2018.07.002], https://hal.inria.fr/hal-01966083

[6] F. BATTISTON, J. GUILLON, M. CHAVEZ, V. LATORA, F. DE VICO FALLANI.*Multiplex core–periphery organization of the human connectome*, in "Journal of the Royal Society Interface", September 2018, vol. 15, n° 146 [*DOI :* 10.1098/RSIF.2018.0514], https://hal.archives-ouvertes.fr/hal-01874871

[7] A. BEAUDET, J. DUMONCEL, F. DE BEER, S. DURRLEMAN, E. GILISSEN, A. OETTLÉ, G. SUBSOL, J. F. THACKERAY, J. BRAGA.*The endocranial shape of Australopithecus africanus : surface analysis of the endocasts of Sts 5 and Sts 60*, in "Journal of Anatomy", February 2018, vol. 232, n° 2, p. 296-303 [*DOI :* 10.1111/JOA.12745], https://hal-lirmm.ccsd.cnrs.fr/lirmm-01636811

[8] A. BERTRAND, J. WEN, D. RINALDI, M. HOUOT, S. SAYAH, A. CAMUZAT, C. FOURNIER, S. FONTANELLA, A. ROUTIER, P. COURATIER, F. PASQUIER, M.-O. HABERT, D. HANNEQUIN, O. MARTINAUD, P. CAROPPO, R. LEVY, B. DUBOIS, A. BRICE, S. DURRLEMAN, O. COLLIOT, I. LE BER, P. STUDY.*Early cognitive, structural and microstructural changes in c9orf72 presymptomatic carriers before 40 years of age*, in "JAMA neurology", February 2018, vol. 75, n° 2, p. 236-245 [*DOI :* 10.1001/JAMANEUROL.2017.4266], https://hal.inria.fr/hal-01654000

[9] S. COLONNESE, M. BIAGI, T. CATTAI, R. CUSANI, F. DE VICO FALLANI, G. SCARANO.*Green Compressive Sampling Reconstruction in IoT Networks*, in "Sensors", August 2018, vol. 18, n° 8, 2735 [*DOI :* 10.3390/S18082735], https://hal.inria.fr/hal-01965543

[10] M.-C. CORSI, M. CHAVEZ, D. SCHWARTZ, L. HUGUEVILLE, A. KHAMBHATI, D. S. BASSETT, F. DE VICO FALLANI.*Integrating EEG and MEG Signals to Improve Motor Imagery Classification in Brain–Computer Interface*, in "International Journal of Neural Systems", April 2018, https://arxiv.org/abs/1711.07258v2 [*DOI :* 10.1142/S0129065718500144], https://hal.archives-ouvertes.fr/hal-01893132

[11] C. CURY, S. DURRLEMAN, D. CASH, M. LORENZI, J. M. NICHOLAS, M. BOCCHETTA, J. C. VAN SWIETEN, B. BORRONI, D. GALIMBERTI, M. MASELLIS, M. C. TARTAGLIA, J. ROWE, C. GRAFF, F. TAGLIAVINI, G. B. FRISONI, R. LAFORCE, E. FINGER, A. DE MENDONÇA, S. SORBI, S. OURSELIN, J. ROHRER, M. MODAT, C. ANDERSSON, S. ARCHETTI, A. ARIGHI, L. BENUSSI, S. BLACK, M. COSSEDDU, M. FALLSTRM, C. G. FERREIRA, C. FENOGLIO, N. FOX, M. FREEDMAN, G. FUMAGALLI, S. GAZZINA, R. GHIDONI, M. GRISOLI, V. JELIC, L. JISKOOT, R. KEREN, G. LOMBARDI, C. MARUTA, L. MEETER, R. VAN MINKELEN, B. NACMIAS, L. IJERSTEDT, A. PADOVANI, J. PANMAN, M. PIEVANI, C. POLITO, E. PREMI, S. PRIONI, R. RADEMAKERS, V. REDAELLI, E. ROGAEVA, G. ROSSI, M. ROSSOR, E. SCARPINI, D. TANG-WAI, H. THONBERG, P. TIRABOSCHI, A. VERDELHO, J. WARREN.*Spatiotemporal analysis for detection of pre-symptomatic shape changes in neurodegenerative diseases: Initial application to the GENFI cohort*, in "NeuroImage", March 2019, vol. 188, p. 282-290 [*DOI :* 10.1016/J.NEUROIMAGE.2018.11.063], https://www.hal.inserm.fr/inserm-01958916

[12] C. CURY, J. GLAUNÈS, R. TORO, M. CHUPIN, G. SCHUMANN, V. FROUIN, J.-B. POLINE, O. COLLIOT.*Statistical Shape Analysis of Large Datasets Based on Diffeomorphic Iterative Centroids*, in "Frontiers

in Neuroscience", November 2018, vol. 12, 803 [*DOI :* 10.3389/FNINS.2018.00803], https://hal.inria.fr/hal-01920263

[13] B. DUBOIS, S. EPELBAUM, F. NYASSE, H. BAKARDJIAN, G. GAGLIARDI, O. USPENSKAYA, M. HOUOT, S. LISTA, F. CACCIAMANI, M.-C. POTIER, A. BERTRAND, F. LAMARI, H. BENALI, J.-F. MANGIN, O. COLLIOT, R. GENTHON, M.-O. HABERT, H. HAMPEL.*Cognitive and neuroimaging features and brain $\beta$-amyloidosis in individuals at risk of Alzheimer's disease (INSIGHT-preAD): a longitudinal observational study*, in "Lancet Neurology", April 2018, vol. 17, n° 4, p. 335 - 346, For the INSIGHT-preAD study group [*DOI :* 10.1016/S1474-4422(18)30029-2], https://hal.sorbonne-universite.fr/hal-01777834

[14] S. EPELBAUM, V. BOUTELOUP, J. F. MANGIN, V. LA CORTE, R. MIGLIACCIO, H. BERTIN, M. O. HABERT, C. FISCHER, C. AZOUANI, L. FILLON, M. CHUPIN, B. VELLAS, F. PASQUIER, J. F. DARTIGUES, F. BLANC, A. GABELLE, M. CECCALDI, P. KROLAK-SALMON, J. HUGON, O. HANON, O. ROUAUD, R. DAVID, G. CHÊNE, B. DUBOIS, C. DUFOUIL.*Neural correlates of episodic memory in the Memento cohort*, in "Alzheimer's & Dementia: Translational Research & Clinical Interventions", 2018, vol. 4, p. 224-233, https://arxiv.org/abs/1812.08031 [*DOI :* 10.1016/J.TRCI.2018.03.010], https://hal.sorbonne-universite.fr/hal-01958834

[15] P. GORI, O. COLLIOT, L. M. KACEM, Y. WORBE, A. ROUTIER, C. POUPON, A. HARTMANN, N. AYACHE, S. DURRLEMAN.*Double diffeomorphism: combining morphometry and structural connectivity analysis*, in "IEEE Transactions on Medical Imaging", September 2018, vol. 37, n° 9, p. 2033-2043 [*DOI :* 10.1109/TMI.2018.2813062], https://hal.archives-ouvertes.fr/hal-01709847

[16] E. HAINQUE, A. BLANCHER, V. MESNAGE, S. RIVAUD-PECHOUX, A. BERTRAND, S. DUPONT, V. NAVARRO, E. ROZE, I. GOURFINKEL-AN, E. APARTIS.*A clinical and neurophysiological motor signature of Unverricht–Lundborg disease*, in "Revue Neurologique", January 2018 [*DOI :* 10.1016/J.NEUROL.2017.06.005], https://hal.archives-ouvertes.fr/hal-01562621

[17] H. HAMPEL, N. TOSCHI, C. BABILONI, F. BALDACCI, K. L. BLACK, A. BOKDE, R. S. BUN, F. CACCIOLA, E. CAVEDO, P. A. CHIESA, O. COLLIOT, C.-M. COMAN, B. DUBOIS, A. DUGGENTO, S. DURRLEMAN, M.-T. FERRETTI, N. GEORGE, R. GENTHON, M.-O. HABERT, K. HERHOLZ, Y. KORONYO, M. KORONYO-HAMAOUI, F. LAMARI, T. LANGEVIN, S. LEHÉRICY, J. LORENCEAU, C. NÉRI, R. NISTICÒ, F. NYASSE, C. W. RITCHIE, S. ROSSI, E. SANTARNECCHI, O. SPORNS, S. R. VERDOONER, A. VERGALLO, N. VILLAIN, E. YOUNESI, F. GARACI, S. LISTA, A. L. BOKDE.*Revolution of Alzheimer Precision Neurology. Passageway of Systems Biology and Neurophysiology*, in "Journal of Alzheimer's Disease", June 2018, vol. 64, n° s1, p. S47 - S105 [*DOI :* 10.3233/JAD-179932], https://hal.archives-ouvertes.fr/hal-01910402

[18] J. P. KIESELMANN, C. P. KAMERLING, N. BURGOS, M. J. MENTEN, C. D. FULLER, S. NILL, M. J. CARDOSO, U. OELFKE.*Geometric and dosimetric evaluations of atlas-based segmentation methods of MR images in the head and neck region*, in "Physics in Medicine and Biology", 2018, vol. 63, n° 14, 00000 [*DOI :* 10.1088/1361-6560/AACB65], https://hal.inria.fr/hal-01827187

[19] I. KOVAL, J.-B. SCHIRATTI, A. M. ROUTIER, M. BACCI, O. COLLIOT, S. ALLASSONNIÈRE, S. DURRLEMAN.*Spatiotemporal Propagation of the Cortical Atrophy: Population and Individual Patterns*, in "Frontiers in Neurology", May 2018, vol. 9, 235 [*DOI :* 10.3389/FNEUR.2018.00235], https://hal.inria.fr/hal-01910400

[20] K. KUMAR, L. CHAUVIN, M. TOEWS, O. COLLIOT, C. DESROSIERS.*Multi-modal brain fingerprinting: a manifold approximation based framework*, in "NeuroImage", December 2018, vol. 183, p. 212 - 226 [*DOI :* 10.1016/J.NEUROIMAGE.2018.08.006], https://hal.inria.fr/hal-01910367

[21] B. LAW-YE, P. DODET, B. HERMANN, S. TRUNET, D. DORMONT, N. PYATIGORSKAYA, D. LECLERCQ.*Progressive white-matter demyelination in delayed CO poisoning encephalopathy*, in "Journal de Neuroradiologie / Journal of Neuroradiology", February 2018, vol. 45, n° 1, p. 59-62 [*DOI :* 10.1016/J.NEURAD.2017.09.008], https://hal.inria.fr/hal-01966086

[22] B. LAW-YE, D. DORMONT, C. CHOLET.*MR angiography of reversible cerebral vasoconstriction syndrome*, in "Diagnostic and Interventional Imaging", September 2018, vol. 99, n° 9, p. 525-526 [*DOI :* 10.1016/J.DIII.2018.07.001], https://hal.inria.fr/hal-01966085

[23] M. LOUIS, B. CHARLIER, P. JUSSELIN, S. PAL, S. DURRLEMAN.*A Fanning Scheme for the Parallel Transport Along Geodesics on Riemannian Manifolds*, in "SIAM Journal on Numerical Analysis", 2018, https://hal.archives-ouvertes.fr/hal-01560787

[24] A. MARCOUX, N. BURGOS, A. BERTRAND, M. TEICHMANN, A. ROUTIER, J. WEN, J. SAMPER-GONZALEZ, S. BOTTANI, S. DURRLEMAN, M.-O. HABERT, O. COLLIOT.*An Automated Pipeline for the Analysis of PET Data on the Cortical Surface*, in "Frontiers in Neuroinformatics", December 2018, vol. 12 [*DOI :* 10.3389/FNINF.2018.00094], https://hal.inria.fr/hal-01950933

[25] J. MARTINEZ, J. BULDÚ, D. PAPO, F. DE VICO FALLANI, M. CHAVEZ.*Role of inter-hemispheric connections in functional brain networks*, in "Scientific Reports", December 2018, vol. 8, n° 1, 10246 [*DOI :* 10.1038/S41598-018-28467-X], https://hal.inria.fr/hal-01964406

[26] A. MENDES, S. TÉZENAS DU MONTCEL, M. LEVY, A. BERTRAND, M.-O. HABERT, H. BERTIN, B. DUBOIS, S. EPELBAUM.*Multimorbidity Is Associated with Preclinical Alzheimer's Disease Neuroimaging Biomarkers*, in "Dementia and Geriatric Cognitive Disorders", August 2018, vol. 45, n° 5-6, p. 272-281 [*DOI :* 10.1159/000489007], https://hal.sorbonne-universite.fr/hal-01958870

[27] C. MEYNIEL, D. SAMRI, F. STEFANO, J. CREVOISIER, F. BONTÉ, R. MIGLIACCIO, L. DELABY, A. BERTRAND, M. O. HABERT, B. DUBOIS, B. BODAGHI, S. EPELBAUM.*COGEVIS: A New Scale to Evaluate Cognition in Patients with Visual Deficiency*, in "Behavioural Neurology", June 2018, vol. 2018, p. 1-7, https://arxiv.org/abs/1812.07328 [*DOI :* 10.1155/2018/4295184], https://hal.sorbonne-universite.fr/hal-01958688

[28] E. PIEKARSKI, N. PYATIGORSKAYA, D. DORMONT, D. GALANAUD, A. KAS.*Increased 18F-FDG Uptake in Lhermitte-Duclos Disease With Cowden Syndrome Revealed by PET-MRI*, in "Clinical Nuclear Medicine", 2018, vol. 43, n° 10, 1 [*DOI :* 10.1097/RLU.0000000000002248], https://hal.inria.fr/hal-01966082

[29] A. ROUTIER, M.-O. HABERT, A. BERTRAND, A. KAS, M. SUNDQVIST, J. MERTZ, P.-M. DAVID, H. BERTIN, S. BELLIARD, F. PASQUIER, K. BENNYS, O. MARTINAUD, F. ETCHARRY-BOUYX, O. MOREAUD, O. GODEFROY, J. PARIENTE, M. PUEL, P. COURATIER, C. BOUTOLEAU-BRETONNIÈRE, B. LAURENT, R. MIGLIACCIO, B. DUBOIS, O. COLLIOT, M. TEICHMANN.*Structural, Microstructural, and Metabolic Alterations in Primary Progressive Aphasia Variants*, in "Frontiers in Neurology", September 2018, vol. 9 [*DOI :* 10.3389/FNEUR.2018.00766], https://hal.inria.fr/hal-01897015

[30] J. SAMPER-GONZALEZ, N. BURGOS, S. BOTTANI, S. FONTANELLA, P. LU, A. MARCOUX, A. ROUTIER, J. GUILLON, M. BACCI, J. WEN, A. BERTRAND, H. BERTIN, M.-O. HABERT, S. DURRLEMAN, T. EVGENIOU, O. COLLIOT.*Reproducible evaluation of classification methods in Alzheimer's disease: Framework and application to MRI and PET data*, in "NeuroImage", December 2018, vol. 183, p. 504-521 [*DOI :* 10.1016/J.NEUROIMAGE.2018.08.042], https://hal.inria.fr/hal-01858384

[31] C. SANCHES, A. ROUTIER, O. COLLIOT, M. TEICHMANN.*The structure of the mental lexicon: what primary progressive aphasias reveal*, in "Neuropsychologia", January 2018, vol. 109, p. 107-115 [*DOI :* 10.1016/J.NEUROPSYCHOLOGIA.2017.12.018], https://hal.inria.fr/hal-01672932

[32] M. SCHERTZ, M. BENZAKOUN, N. PYATIGORSKAYA, S. BELKACEM, M. SAHLI-AMOR, V. NAVARRO, C. CHOLET, D. LECLERCQ, D. DORMONT, B. LAW-YE.*Specificities of arterial spin labeling (ASL) abnormalities in acute seizure*, in "Journal de Neuroradiologie / Journal of Neuroradiology", November 2018 [*DOI :* 10.1016/J.NEURAD.2018.11.003], https://hal.inria.fr/hal-01966079

[33] C. SCOTT, J. JIAO, A. MELBOURNE, N. BURGOS, D. CASH, E. DE VITA, P. MARKIEWICZ, A. O'CONNOR, D. THOMAS, P. S. WESTON, J. SCHOTT, B. HUTTON, S. OURSELIN.*Reduced acquisition time PET pharmacokinetic modelling using simultaneous ASL–MRI: proof of concept*, in "Journal of Cerebral Blood Flow and Metabolism", September 2018 [*DOI :* 10.1177/0271678X18797343], https://hal.inria.fr/hal-01871983

[34] J. WEN, H. ZHANG, D. C. ALEXANDER, S. DURRLEMAN, A. ROUTIER, D. RINALDI, M. HOUOT, P. COURATIER, D. HANNEQUIN, F. PASQUIER, J. ZHANG, O. COLLIOT, I. LE BER, A. BERTRAND.*Neurite density is reduced in the presymptomatic phase of C9orf72 disease*, in "Journal of Neurology, Neurosurgery and Psychiatry", 2018 [*DOI :* 10.1136/JNNP-2018-318994], https://hal.inria.fr/hal-01907482

[35] C. ZAVANONE, Y. SAMSON, C. ARBIZU, S. DUPONT, D. DORMONT, C. ROSSO.*Critical brain regions related to post-stroke aphasia severity identified by early diffusion imaging are not the same when predicting short- and long-term outcome*, in "Brain and Language", November 2018, vol. 186, p. 1-7 [*DOI :* 10.1016/J.BANDL.2018.08.005], https://hal.inria.fr/hal-01966081

### Invited Conferences

[36] A. BÔNE, M. LOUIS, B. MARTIN, S. DURRLEMAN.*Deformetrica 4: an open-source software for statistical shape analysis*, in "ShapeMI @ MICCAI 2018 - Workshop on Shape in Medical Imaging", Granada, Spain, September 2018, https://hal.inria.fr/hal-01874752

### International Conferences with Proceedings

[37] M. ANSART, I. KOVAL, A. BERTRAND, D. DORMONT, S. DURRLEMAN.*Design of a decision support system for predicting the progression of Alzheimer's disease*, in "Alzheimer's Association International Conference", Chicago, United States, July 2018, vol. 14, n$^{o}$ 7, P433 [*DOI :* 10.1016/J.JALZ.2018.06.371], https://hal.inria.fr/hal-01988912

[38] A. BÔNE, O. COLLIOT, S. DURRLEMAN.*Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms*, in "CVPR 2018 - Computer Vision and Pattern Recognition 2018", Salt Lake City, United States, June 2018, https://hal.archives-ouvertes.fr/hal-01744538

[39] V. J. HENRY, I. MOSZER, O. DAMERON, M.-C. POTIER, M. HOFMANN-APITIUS, O. COLLIOT.*Converting Alzheimer's disease map into a heavyweight ontology: a formal network to integrate data*, in "DILS 2018 - 13th International Conference on Data Integration in the Life Sciences", Hannover, Germany, DILS 2018 - 13th International Conference on Data Integration in the Life Sciences, November 2018, p. 1-9, https://hal.archives-ouvertes.fr/hal-01917742

[40] W. WEI, E. POIRION, B. BODINI, S. DURRLEMAN, N. AYACHE, B. STANKOFF, O. COLLIOT.*Learning Myelin Content in Multiple Sclerosis from Multimodal MRI through Adversarial Training*, in "MICCAI 2018 – 21st International Conference On Medical Image Computing & Computer Assisted Intervention", Granada, Spain, September 2018, vol. 11072 [*DOI :* 10.1007/978-3-030-00931-1_59], https://hal.inria.fr/hal-01810822

[41] J. WEN, J. SAMPER-GONZALEZ, S. BOTTANI, A. ROUTIER, N. BURGOS, T. JACQUEMONT, S. FONTANELLA, S. DURRLEMAN, A. BERTRAND, O. COLLIOT.*Using diffusion MRI for classification and prediction of Alzheimer's Disease: a reproducible study*, in "AAIC 2018 - Alzheimer's Association International Conference", Chicago, United States, July 2018, https://hal.inria.fr/hal-01758167

### Conferences without Proceedings

[42] G. BASSIGNANA, J. FRANSSON, O. COLLIOT, V. ZUJOVIC, F. DE VICO FALLANI.*Identification of Driver Nodes in Genetic Networks Regulating Macrophage Activation*, in "Conference on Complex Systems - CSS", Thessaloniki, Greece, September 2018, https://hal.archives-ouvertes.fr/hal-01967797

[43] M.-C. CORSI, M. CHAVEZ, D. SCHWARTZ, L. HUGUEVILLE, A. KHAMBHATI, D. S. BASSETT, F. DE VICO FALLANI.*Integrating EEG and MEG information to enhance motor-imagery classification in brain-computer interface*, in "BIOMAG 2018 - 21st International Conference on Biomagnetism", Philadelphia, United States, August 2018, https://hal.archives-ouvertes.fr/hal-01966311

[44] M.-C. CORSI, M. CHAVEZ, D. SCHWARTZ, L. HUGUEVILLE, A. KHAMBHATI, D. S. BASSETT, F. DE VICO FALLANI.*M/EEG integration to enhance motor-imagery-based brain-computer interface performances*, in "Seventh International BCI Society Meeting", Asilomar, United States, May 2018, https://hal.archives-ouvertes.fr/hal-01966314

[45] V. HENRY, I. MOSZER, O. DAMERON, M.-C. POTIER, M. HOFMANN-APITIUS, O. COLLIOT.*Integrating ontological representation and reasoning into a disease map: application to Alzheimer's disease*, in "DMCM 2018 - 3rd Disease Maps Community Meeting", Paris, France, June 2018, p. 1-2, https://hal.archives-ouvertes.fr/hal-01873474

[46] A. MARCOUX, N. BURGOS, A. BERTRAND, A. ROUTIER, J. WEN, J. SAMPER-GONZALEZ, S. BOTTANI, S. DURRLEMAN, M.-O. HABERT, O. COLLIOT.*A pipeline for the analysis of 18F-FDG PET data on the cortical surface and its evaluation on ADNI*, in "Annual meeting of the Organization for Human Brain Mapping - OHBM 2018", Singapour, Singapore, June 2018, https://hal.archives-ouvertes.fr/hal-01757646

[47] A. ROUTIER, J. GUILLON, N. BURGOS, J. SAMPER-GONZALEZ, J. WEN, S. FONTANELLA, S. BOTTANI, T. JACQUEMONT, A. MARCOUX, P. GORI, P. LU, T. MOREAU, M. BACCI, S. DURRLEMAN, O. COLLIOT.*Clinica: an open source software platform for reproducible clinical neuroscience studies*, in "Annual meeting of the Organization for Human Brain Mapping - OHBM 2018", Singapore, Singapore, June 2018, https://hal.inria.fr/hal-01760658

[48] A. ROUTIER, M.-O. HABERT, A. BERTRAND, A. KAS, P.-M. DAVID, H. BERTIN, O. GODEFROY, F. ETCHARRY-BOUYX, O. MOREAUD, F. PASQUIER, P. COURATIER, K. BENNYS, C. BOUTOLEAU-BRETONNIÈRE, O. MARTINAUD, B. LAURENT, J. PARIENTE, M. PUEL, S. BELLIARD, R. MIGLIACCIO, B. DUBOIS, O. COLLIOT, M. TEICHMANN.*Structural, microstructural and metabolic alterations in Primary Progressive Aphasia variants*, in "Annual meeting of the Organization for Human Brain Mapping - OHBM 2018", Singapore, Singapore, June 2018, https://hal.inria.fr/hal-01764289

[49] J. SAMPER-GONZALEZ, S. BOTTANI, N. BURGOS, S. FONTANELLA, P. LU, A. MARCOUX, A. ROUTIER, J. GUILLON, M. BACCI, J. WEN, A. BERTRAND, H. BERTIN, M.-O. HABERT, S. DURRLEMAN, T. EVGENIOU, O. COLLIOT.*Reproducible evaluation of Alzheimer's Disease classification from MRI and PET data*, in "Annual meeting of the Organization for Human Brain Mapping - OHBM 2018", Singapour, Singapore, June 2018, https://hal.inria.fr/hal-01761666

[50] W. WEI, E. POIRION, B. BODINI, S. DURRLEMAN, O. COLLIOT, B. STANKOFF, N. AYACHE.*FLAIR MR Image Synthesis By Using 3D Fully Convolutional Networks for Multiple Sclerosis*, in "ISMRM-ESMRMB 2018 - Joint Annual Meeting", Paris, France, June 2018, p. 1-6, https://hal.inria.fr/hal-01723070

[51] J. WEN, J. SAMPER-GONZALEZ, S. BOTTANI, A. ROUTIER, N. BURGOS, T. JACQUEMONT, S. FONTANELLA, S. DURRLEMAN, A. BERTRAND, O. COLLIOT.*Comparison of DTI Features for the Classification of Alzheimer's Disease: A Reproducible Study*, in "OHBM 2018 - Organization for Human Brain Mapping Annual Meeting", Singapour, Singapore, June 2018, https://hal.inria.fr/hal-01758206

[52] J. WEN, H. ZHANG, D. ALEXANDER, S. DURRLEMAN, A. ROUTIER, D. RINALDI, M. HOUOT, J. ZHANG, O. COLLIOT, I. LE BER, A. BERTRAND.*NODDI Highlights Promising New Markers In Presymptomatic C9orf72 Carriers*, in "OHBM 2018 - Organization for Human Brain Mapping Annual Meeting", Singapour, Singapore, June 2018, https://hal.inria.fr/hal-01758137

## Other Publications

[53] A. BÔNE, M. LOUIS, O. COLLIOT, S. DURRLEMAN.*Learning low-dimensional representations of shape data sets with diffeomorphic autoencoders*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01963736

[54] C. CURY, S. DURRLEMAN, D. M. CASH, M. LORENZI, J. M. NICHOLAS, M. BOCCHETTA, J. C. VAN SWIETEN, B. BORRONI, D. GALIMBERTI, M. MASELLIS, M. C. TARTAGLIA, J. ROWE, C. GRAFF, F. TAGLIAVINI, G. B. FRISONI, R. J. LAFORCE, E. FINGER, A. DE MENDONÇA, S. SORBI, S. OURSELIN, J. D. ROHRER, M. M. MODAT.*Spatiotemporal analysis for detection of pre-symptomatic shape changes in neurodegenerative diseases: applied to GENFI study*, August 2018, working paper or preprint [*DOI :* 10.1101/385427], https://hal.inria.fr/hal-01856906

[55] C. CURY, J. A. GLAUNÈS, R. TORO, M. CHUPIN, G. SCHUMANN, V. FROUIN, J. B. POLINE, O. COLLIOT.*Statistical shape analysis of large datasets based on diffeomorphic iterative centroids*, July 2018, working paper or preprint [*DOI :* 10.1101/363861], https://hal.inria.fr/hal-01832191

[56] I. KOVAL, A. BÔNE, M. LOUIS, S. BOTTANI, A. MARCOUX, J. SAMPER-GONZALEZ, N. BURGOS, B. CHARLIER, A. BERTRAND, S. EPELBAUM, O. COLLIOT, S. ALLASSONNIÈRE, S. DURRLEMAN.*Simulating Alzheimer's disease progression with personalised digital brain models*, 2018, working paper or preprint, https://hal.inria.fr/hal-01964821

[57] M. LOUIS, B. CHARLIER, S. DURRLEMAN.*Learning Riemannian geometry for mixed-effect models using deep generative networks*, July 2018, working paper or preprint, https://hal.inria.fr/hal-01828949

[58] C. OBANDO FORERO, F. DE VICO FALLANI.*Temporal metrics for exponential random graph models of time-evolving networks*, June 2018, NetSci 2018 - International School and Conference in Complex Science, https://hal.inria.fr/hal-01967589

# Project-Team CAGE

## Control and Geometry

IN COLLABORATION WITH: Laboratoire Jacques-Louis Lions (LJLL)

IN PARTNERSHIP WITH:
**CNRS**

**Sorbonne Université**

RESEARCH CENTER
**Paris**

THEME
**Optimization and control of dynamic systems**

# Table of contents

# Project-Team CAGE

*Creation of the Team: 2017 July 01, updated into Project-Team: 2018 August 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)
A6.4. - Automatic control
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization
A6.4.5. - Control of distributed parameter systems
A6.4.6. - Optimal control

**Other Research Topics and Application Domains:**

B1.2. - Neuroscience and cognitive science
B2.6. - Biological and medical imaging
B5.11. - Quantum systems
B7.1.3. - Air traffic

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Mario Sigalotti [Team leader, Inria, Senior Researcher, HDR]
Ugo Boscain [CNRS, Senior Researcher, HDR]

**Faculty Members**

Jean Michel Coron [Sorbonne Université, Professor, HDR]
Emmanuel Trélat [Sorbonne Université, Professor, HDR]

**PhD Students**

Nicolas Augier [Ecole polytechnique]
Riccardo Bonalli [ONERA, until Sep 2018]
Amaury Hayat [Sorbonne Université, from Oct 2016]
Mathieu Kohli [Ecole polytechnique]
Gontran Lance [Sorbonne Université]
Cyril Letrouit [Ecole Normale Supérieure Paris, from Jun 2018]
Antoine Olivier [Univ Pierre et Marie Curie, until Oct 2018]
Jakub Orlowski [Centrale-Supélec]
Eugenio Pozzoli [Inria, from Sep 2018]
Ludovic Sacchelli [Ecole polytechnique, until Oct 2018]
Shengquan Xiang [Univ Pierre et Marie Curie]
Christophe Zhang [Univ Pierre et Marie Curie]

**Post-Doctoral Fellows**

Ivan Beschastnyi [ANR SRGI, from Oct 2018]
Francesco Boarotto [Ecole polytechnique, until Jun 2018]
Valentina Franceschi [Inria, until Mar 2018]

**Administrative Assistant**
Mathieu Mourey [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

CAGE's activities take place in the field of mathematical control theory, with applications in three main directions: geometric models for vision, control of quantum mechanical systems, and control of systems with uncertain dynamics.

The relations between control theory and geometry of vision rely on the notion of sub-Riemannian structure, a geometric framework which is used to measure distances in nonholonomic contexts and which has a natural and powerful control theoretical interpretation. We recall that nonholonomicity refers to the property of a velocity constraint that cannot be recast as a state constraint. In the language of differential geometry, a sub-Riemannian structure is a (possibly rank-varying) Lie bracket generating distribution endowed with a smoothly varying norm.

Sub-Riemannian geometry, and in particular the theory of associated (hypoelliptic) diffusive processes, plays a crucial role in the neurogeometrical model of the primary visual cortex due to Petitot, Citti and Sarti, based on the functional architecture first described by Hubel and Wiesel. Such a model can be used as a powerful paradigm for bio-inspired image processing, as already illustrated in the recent literature (including by members of our team). Our contributions to this field are based not only on this approach, but also on another geometric and sub-Riemannian framework for vision, based on pattern matching in the group of diffeomorphisms. In this case admissible diffeomorphisms correspond to deformations which are generated by vector fields satisfying a set of nonholonomic constraints. A sub-Riemannian metric on the infinite-dimensional group of diffeomorphisms is induced by a length on the tangent distribution of admissible velocities. Nonholonomic constraints can be especially useful to describe distortions of sets of interconnected objects (e.g., motions of organs in medical imaging).

Control theory is one of the components of the forthcoming quantum revolution [0], since manipulation of quantum mechanical systems is ubiquitous in applications such as quantum computation, quantum cryptography, and quantum sensing (in particular, imaging by nuclear magnetic resonance). The efficiency of the control action has a dramatic impact on the quality of the coherence and the robustness of the required manipulation. Minimal time constraints and interaction of time scales are important factors for characterizing the efficiency of a quantum control strategy. Time scales analysis is important for evaluation approaches based on adiabatic approximation theory, which is well-known to improve the robustness of the control strategy. CAGE works for the improvement of evaluation and design tools for efficient quantum control paradigms, especially for what concerns quantum systems evolving in infinite-dimensional Hilbert spaces.

Simultaneous control of a continuum of systems with slightly different dynamics is a typical problem in quantum mechanics and also a special case of the third applicative axis to which CAGE is contributing: control of systems with uncertain dynamics. The slightly different dynamics can indeed be seen as uncertainties in the system to be controlled, and simultaneous control rephrased in terms of a robustness task. Robustification, i.e., offsetting uncertainties by suitably designing the control strategy, is a widespread task in automatic control theory, showing up in many applicative domains such as electric circuits or aerospace motion planning. If dynamics are not only subject to static uncertainty, but may also change as time goes, the problem of controlling the system can be recast within the theory of switched and hybrid systems, both in a deterministic and in a probabilistic setting. Our contributions to this research field concern both stabilization (either asymptotic or in finite time) and optimal control, where redundancies and probabilistic tools can be introduced to offset uncertainties.

[0] As anticipated by the recent launch of the FET Flagship on Quantum Technologies

# 3. Research Program

## 3.1. Research domain

The activities of CAGE are part of the research in the wide area of control theory. This nowadays mature discipline is still the subject of intensive research because of its crucial role in a vast array of applications.

More specifically, our contributions are in the area of **mathematical control theory**, which is to say that we are interested in the analytical and geometrical aspects of control applications. In this approach, a control system is modeled by a system of equations (of many possible types: ordinary differential equations, partial differential equations, stochastic differential equations, difference equations,...), possibly not explicitly known in all its components, which are studied in order to establish qualitative and quantitative properties concerning the actuation of the system through the control.

**Motion planning** is, in this respect, a cornerstone property: it denotes the design and validation of algorithms for identifying a control law steering the system from a given initial state to (or close to) a target one. Initial and target positions can be replaced by sets of admissible initial and final states as, for instance, in the motion planning task towards a desired periodic solution. Many specifications can be added to the pure motion planning task, such as robustness to external or endogenous disturbances, obstacle avoidance or penalization criteria. A more abstract notion is that of **controllability**, which denotes the property of a system for which any two states can be connected by a trajectory corresponding to an admissible control law. In mathematical terms, this translates into the surjectivity of the so-called **end-point map**, which associates with a control and an initial state the final point of the corresponding trajectory. The analytical and topological properties of endpoint maps are therefore crucial in analyzing the properties of control systems.

One of the most important additional objective which can be associated with a motion planning task is **optimal control**, which corresponds to the minimization of a cost (or, equivalently, the maximization of a gain) [156]. Optimal control theory is clearly deeply interconnected with calculus of variations, even if the non-interchangeable nature of the time-variable results in some important specific features, such as the occurrence of **abnormal extremals** [119]. Research in optimal control encompasses different aspects, from numerical methods to dynamic programming and non-smooth analysis, from regularity of minimizers to high order optimality conditions and curvature-like invariants.

Another domain of control theory with countless applications is **stabilization**. The goal in this case is to make the system converge towards an equilibrium or some more general safety region. The main difference with respect to motion planning is that here the control law is constructed in feedback form. One of the most important properties in this context is that of **robustness**, i.e., the performance of the stabilization protocol in presence of disturbances or modeling uncertainties. A powerful framework which has been developed to take into account uncertainties and exogenous non-autonomous disturbances is that of hybrid and switched systems [159], [118], [147]. The central tool in the stability analysis of control systems is that of **control Lyapunov function**. Other relevant techniques are based on algebraic criteria or dynamical systems. One of the most important stability property which is studied in the context of control system is **input-to-state stability** [143], which measures how sensitive the system is to an external excitation.

One of the areas where control applications have nowadays the most impressive developments is in the field of **biomedicine and neurosciences**. Improvements both in modeling and in the capability of finely actuating biological systems have concurred in increasing the popularity of these subjects. Notable advances concern, in particular, identification and control for biochemical networks [137] and models for neural activity [105]. Therapy analysis from the point of view of optimal control has also attracted a great attention [140].

Biological models are not the only one in which stochastic processes play an important role. Stock-markets and energy grids are two major examples where optimal control techniques are applied in the non-deterministic setting. Sophisticated mathematical tools have been developed since several decades to allow for such extensions. Many theoretical advances have also been required for dealing with complex systems whose description is based on **distributed parameters** representation and **partial differential equations**. Functional analysis, in particular, is a crucial tool to tackle the control of such systems [153].

Let us conclude this section by mentioning another challenging application domain for control theory: the decision by the European Union to fund a flagship devoted to the development of quantum technologies is a symptom of the role that quantum applications are going to play in tomorrow's society. **Quantum control** is one of the bricks of quantum engineering, and presents many peculiarities with respect to standard control theory, as a consequence of the specific properties of the systems described by the laws of quantum physics. Particularly important for technological applications is the capability of inducing and reproducing coherent state superpositions and entanglement in a fast, reliable, and efficient way [106].

## 3.2. Scientific foundations

At the core of the scientific activity of the team is the **geometric control** approach, that is, a distinctive viewpoint issued in particular from (elementary) differential geometry, to tackle questions of controllability, observability, optimal control... [68], [110]. The emphasis of such a geometric approach to control theory is put on intrinsic properties of the systems and it is particularly well adapted to study nonlinear and nonholonomic phenomena.

One of the features of the geometric control approach is its capability of exploiting **symmetries and intrinsic structures** of control systems. Symmetries and intrinsic structures can be used to characterize minimizing trajectories, prove regularity properties and describe invariants. An egregious example is given by mechanical systems, which inherently exhibit Lagrangian/Hamiltonian structures which are naturally expressed using the language of symplectic geometry [91]. The geometric theory of quantum control, in particular, exploits the rich geometric structure encoded in the Schrödinger equation to engineer adapted control schemes and to characterize their qualitative properties. The Lie–Galerkin technique that we proposed starting from 2009 [94] builds on this premises in order to provide powerful tests for the controllability of quantum systems defined on infinite-dimensional Hilbert spaces.

Although the focus of geometric control theory is on qualitative properties, its impact can also be disruptive when it is used in combination with quantitative analytical tools, in which case it can dramatically improve the computational efficiency. This is the case in particular in optimal control. Classical optimal control techniques (in particular, Pontryagin Maximum Principle, conjugate point theory, associated numerical methods) can be significantly improved by combining them with powerful modern techniques of geometric optimal control, of the theory of numerical continuation, or of dynamical system theory [152], [139]. Geometric optimal control allows the development of general techniques, applying to wide classes of nonlinear optimal control problems, that can be used to characterize the behavior of optimal trajectories and in particular to establish regularity properties for them and for the cost function. Hence, geometric optimal control can be used to obtain powerful optimal syntheses results and to provide deep geometric insights into many applied problems. Numerical optimal control methods with geometric insight are in particular important to handle subtle situations such as rigid optimal paths and, more generally, optimal syntheses exhibiting abnormal minimizers.

Optimal control is not the only area where the geometric approach has a great impact. Let us mention, for instance, motion planning, where different geometric approaches have been developed: those based on the **Lie algebra** associated with the control system [132], [121], those based on the differentiation of nonlinear flows such as the **return method** [99], [98], and those exploiting the **differential flatness** of the system [103].

Geometric control theory is not only a powerful framework to investigate control systems, but also a useful tool to model and study phenomena that are not *a priori* control-related. Two occurrences of this property play an important role in the activities of CAGE:

- geometric control theory as a tool to investigate properties of mathematical structures;
- geometric control theory as a modeling tool for neurophysical phenomena and for synthesizing biomimetic algorithms based on such models.

Examples of the first type, concern, for instance, hypoelliptic heat kernels [66] or shape optimization [74]. Examples of the second type are inactivation principles in human motricity [77] or neurogeometrical models for image representation of the primary visual cortex in mammals [88].

A particularly relevant class of control systems, both from the point of view of theory and applications, is characterized by the linearity of the controlled vector field with respect to the control parameters. When the controls are unconstrained in norm, this means that the admissible velocities form a distribution in the tangent bundle to the state manifold. If the distribution is equipped with a point-dependent quadratic form (encoding the cost of the control), the resulting geometrical structure is said to be **sub-Riemannian**. Sub-Riemannian geometry appears as the underlying geometry of nonlinear control systems: in a similar way as the linearization of a control system provides local informations which are readable using the Euclidean metric scale, sub-Riemannian geometry provides an adapted non-isotropic class of lenses which are often much more informative. As such, its study is fundamental for control design. The importance of sub-Riemannian geometry goes beyond control theory and it is an active field of research both in differential geometry [129], geometric measure theory [70] and hypoelliptic operator theory [80].

The geometric control approach has historically been related to the development of finite-dimensional control theory. However, its impact in the analysis of distributed parameter control systems and in particular systems of controlled partial differential equations has been growing in the last decades, complementing analytical and numerical approaches, providing dynamical, qualitative and intrinsic insight [97]. CAGE's ambition is to be at the core of this development in the years to come.

# 4. Application Domains

## 4.1. First axis: Geometry of vision

A suggestive application of sub-Riemannian geometry and in particular of hypoelliptic diffusion comes from a model of geometry of vision describing the functional architecture of the primary visual cortex V1. In 1958, Hubel and Wiesel (Nobel in 1981) observed that the visual cortex V1 is endowed with the so-called **pinwheel structure**, characterized by neurons grouped into orientation columns, that are sensible both to positions and directions [109]. The mathematical rephrasing of this discovery is that the visual cortex lifts an image from $\mathbf{R}^2$ into the bundle of directions of the plane [95], [136], [138], [108].

A simplified version of the model can be described as follows: neurons of V1 are grouped into orientation columns, each of them being sensitive to visual stimuli at a given point of the retina and for a given direction on it. The retina is modeled by the real plane, i.e., each point is represented by a pair $(x, y) \in \mathbb{R}^2$, while the directions at a given point are modeled by the projective line, i.e. an element $\theta$ of the projective line $P^1$. Hence, the primary visual cortex V1 is modeled by the so called projective tangent bundle $\mathrm{PTR}^2 = \mathbf{R}^2 \times \mathbf{P}^1$. From a neurological point of view, orientation columns are in turn grouped into hypercolumns, each of them being sensitive to stimuli at a given point $(x, y)$ with any direction.

Orientation columns are connected between them in two different ways. The first kind of connections are the vertical (inhibitory) ones, which connect orientation columns belonging to the same hypercolumn and sensible to similar directions. The second kind of connections are the horizontal (excitatory) connections, which connect neurons belonging to different (but not too far) hypercolumns and sensible to the same directions. The resulting metric structure is sub-Riemannian and the model obtained in this way provides a convincing explanation in terms of sub-Riemannian geodesics of gestalt phenomena such as Kanizsa illusory contours.

The sub-Riemannian model for image representation of V1 has a great potential of yielding powerful bio-inspired image processing algorithms [102], [88]. Image inpainting, for instance, can be implemented by reconstructing an incomplete image by activating orientation columns in the missing regions in accordance with sub-Riemannian non-isotropic constraints. The process intrinsically defines an hypoelliptic heat equation on $\mathrm{PTR}^2$ which can be integrated numerically using non-commutative Fourier analysis on a suitable semidiscretization of the group of roto-translations of the plane [86].

We have been working on the model and its software implementation since 2012. This work has been supported by several project, as the ERC starting grant GeCoMethods and the ERC Proof of Concept ARTIV1 of U. Boscain, and the ANR GCM.

A parallel approach that we will pursue and combine with this first one is based on **pattern matching in the group of diffeomorphisms**. We want to extend this approach, already explored in the Riemannian setting [151], [126], to the general sub-Riemannian framework. The paradigm of the approach is the following: consider a distortable object, more or less rigid, discretized into a certain number of points. One may track its distortion by considering the paths drawn by these points. One would however like to know how the object itself (and not its discretized version) has been distorted. The study in [151], [126] shed light on the importance of Riemannian geometry in this kind of problem. In particular, they study the Riemannian submersion obtained by making the group of diffeomorphisms act transitively on the manifold formed by the points of the discretization, minimizing a certain energy so as to take into account the whole object. Settled as such, the problem is Riemannian, but if one considers objects involving connections, or submitted to nonholonomic constraints, like in medical imaging where one tracks the motions of organs, then one comes up with a sub-Riemannian problem. The transitive group is then far bigger, and the aim is to lift curves submitted to these nonholonomic constraints into curves in the set of diffeomorphisms satisfying the corresponding constraints, in a unique way and minimizing an energy (giving rise to a sub-Riemannian structure).

## 4.2. Second axis: Quantum control

The goal of quantum control is to design efficient protocols for tuning the occupation probabilities of the energy levels of a system. This task is crucial in atomic and molecular physics, with applications ranging from photochemistry to nuclear magnetic resonance and quantum computing. A quantum system may be controlled by exciting it with one or several external fields, such as magnetic or electric fields. The goal of quantum control theory is to adapt the tools originally developed by control theory and to develop new specific strategies that tackle and exploit the features of quantum dynamics (probabilistic nature of wavefunctions and density operators, measure and wavefunction collapse, decoherence, ...). A rich variety of relevant models for controlled quantum dynamics exist, encompassing low-dimensional models (e.g., single-spin systems) and PDEs alike, with deterministic and stochastic components, making it a rich and exciting area of research in control theory.

The controllability of quantum system is a well-established topic when the state space is finite-dimensional [100], thanks to general controllability methods for left-invariant control systems on compact Lie groups [90], [111]. When the state space is infinite-dimensional, it is known that in general the bilinear Schrödinger equation is not exactly controllable [154]. Nevertheless, weaker controllability properties, such as approximate controllability or controllability between eigenstates of the internal Hamiltonian (which are the most relevant physical states), may hold. In certain cases, when the state space is a function space on a 1D manifold, some rather precise description of the set of reachable states has been provided [75]. A similar description for higher-dimensional manifolds seems intractable and at the moment only approximate controllability results are available [127], [134], [112]. The most widely applicable tests for controllability of quantum systems in infinite-dimensional Hilbert spaces are based on the **Lie–Galerkin technique** [94], [83], [84]. They allow, in particular, to show that the controllability property is generic among this class of systems [124].

A family of algorithms which are specific to quantum systems are those based on adiabatic evolution [158], [157], [115]. The basic principle of adiabatic control is that the flow of a slowly varying Hamiltonian can be approximated (up to a phase factor) by a quasi-static evolution, with a precision proportional to the velocity of variation of the Hamiltonian. The advantage of the **adiabatic approach** is that it is constructive and produces control laws which are both smooth and robust to parameter uncertainty. The paradigm is based on the adiabatic perturbation theory developed in mathematical physics [81], [133], [150], where it plays an important role for understanding molecular dynamics. Approximation theory by adiabatic perturbation can be used to describe the evolution of the occupation probabilities of the energy levels of a slowly varying Hamiltonian. Results from the last 15 years, including those by members of our team [62], [87], have highlighted the effectiveness of control techniques based on adiabatic path following.

## 4.3. Third axis: Stability and uncertain dynamics

Switched and hybrid systems constitute a broad framework for the description of the heterogeneous aspects of systems in which continuous dynamics (typically pertaining to physical quantities) interact with discrete/logical components. The development of the switched and hybrid paradigm has been motivated by a broad range of applications, including automotive and transportation industry [142], energy management [135] and congestion control [125].

Even if both controllability [146] and observability [113] of switched and hybrid systems have attracted much research efforts, the central role in their study is played by the problem of stability and stabilizability. The goal is to determine whether a dynamical or a control system whose evolution is influenced by a time-dependent signal is uniformly stable or can be uniformly stabilized [118], [147]. Uniformity is considered with respect to all signals in a given class. Stability of switched systems lead to several interesting phenomena. For example, even when all the subsystems corresponding to a constant switching law are exponentially stable, the switched systems may have divergent trajectories for certain switching signals [117]. This fact illustrates the fact that stability of switched systems depends not only on the dynamics of each subsystem but also on the properties of the class of switching signals which is considered.

The most common class of switching signals which has been considered in the literature is made of all piecewise constant signals. In this case uniform stability of the system is equivalent to the existence of a common quadratic Lyapunov function [128]. Moreover, provided that the system has finitely many modes, the Lyapunov function can be taken polyhedral or polynomial [78], [79], [101]. A special role in the switched control literature has been played by common quadratic Lyapunov functions, since their existence can be tested rather efficiently (see the surveys [120], [141] and the references therein). It is known, however, that the existence of a common quadratic Lyapunov function is not necessary for the global uniform exponential stability of a linear switched system with finitely many modes. Moreover, there exists no uniform upper bound on the minimal degree of a common polynomial Lyapunov function [123]. More refined tools rely on multiple and non-monotone Lyapunov functions [89]. Let us also mention linear switched systems technics based on the analysis of the Lie algebra generated by the matrices corresponding to the modes of the system [65].

For systems evolving in the plane, more geometrical tests apply, and yield a complete characterization of the stability [82], [71]. Such a geometric approach also yields sufficient conditions for uniform stability in the linear planar case [85].

In many situations, it is interesting for modeling purposes to specify the features of the switched system by introducing **constrained switching rules**. A typical constraint is that each mode is activated for at least a fixed minimal amount of time, called the dwell-time. Switching rules can also be imposed, for instance, by a timed automata. When constraints apply, the common Lyapunov function approach becomes conservative and new tools have to be developed to give more detailed characterizations of stable and unstable systems.

Our approach to constrained switching is based on the idea of relating the analytical properties of the classes of constrained switching laws (shift-invariance, compactness, closure under concatenation, ...) to the stability behavior of the corresponding switched systems. One can introduce **probabilistic uncertainties** by endowing the classes of admissible signals with suitable probability measures. One then looks at the corresponding Lyapunov exponents, whose existence is established by the multiplicative ergodic theorem. The interest of this approach is that probabilistic stability analysis filters out highly 'exceptional' worst-case trajectories. Although less explicitly characterized from a dynamical viewpoint than its deterministic counterpart, the probabilistic notion of uniform exponential stability can be studied using several reformulations of Lyapunov exponents proposed in the literature [76], [96], [155].

## 4.4. Joint theoretical core

The theoretical questions raised by the different applicative area will be pooled in a research axis on the transversal aspects of geometric control theory and sub-Riemannian structures.

We recall that sub-Riemannian geometry is a generalization of Riemannian geometry, whose birth dates back to Carathéodory's seminal paper on the foundations of Carnot thermodynamics [92], followed by E. Cartan's address at the International Congress of Mathematicians in Bologna [93]. In the last twenty years,

sub-Riemannian geometry has emerged as an independent research domain, with a variety of motivations and ramifications in several parts of pure and applied mathematics. Let us mention geometric analysis, geometric measure theory, stochastic calculus and evolution equations together with applications in mechanics and optimal control (motion planning, robotics, nonholonomic mechanics, quantum control) [60], [61].

One of the main open problems in sub-Riemannian geometry concerns the regularity of length-minimizers [63], [130]. Length-minimizers are solutions to a variational problem with constraints and satisfy a first-order necessary condition resulting from the Pontryagin Maximum Principle (PMP). Solutions of the PMP are either *normal* or *abnormal*. Normal length-minimizer are well-known to be smooth, i.e., $C^\infty$, as it follows by the Hamiltonian nature of the PMP. The question of regularity is then reduced to abnormal length-minimizers. If the sub-Riemannian structure has step 2, then abnormal length-minimizers can be excluded and thus every length-minimizer is smooth. For step 3 structures, the situation is already more complicated and smoothness of length-minimizers is known only for Carnot groups [114], [149]. The question of regularity of length-minimizers is not restricted to the smoothness in the $C^\infty$ sense. A recent result prove that length-minimizers, for sub-Riemannian structures of any step, cannot have corner-like singularities [107]. When the sub-Riemannian structure is analytic, more is known on the size of the set of points where a length-minimizer can lose analyticity [148], regardless of the rank and of the step of the distribution.

An interesting set of recent results in sub-Riemannian geometry concerns the extension to such a setting of the Riemannian notion of sectional curvature. The curvature operator can be introduced in terms of the symplectic invariants of the Jacobi curve [67], [116], [64], a curve in the Lagrange Grassmannian related to the linearization of the Hamiltonian flow. Alternative approaches to curvatures in metric spaces are based either on the associated heat equation and the generalization of the curvature-dimension inequality [72], [73] or on optimal transport and the generalization of Ricci curvature [145], [144], [122], [69].

# 5. Highlights of the Year

## 5.1. Highlights of the Year

**Emmanuel Trélat** has been invited speaker at the International Congress of Mathematicians (ICM2018) in Rio, Brazil, in the session “Control theory and optimization".

### *5.1.1. Awards*

- The poster “Adaptive Stimulation Strategy for Selective Brain Oscillations Disruption in a Neuronal Population Model with Delays" by **Jakub Orlowski**, Antoine Chaillet, **Mario Sigalotti**, and Alain Destexhe, has received the CPHS 2018 Best Poster Prize at the 2nd IFAC Conference on Cyber-Physical & Human Systems.

# 6. New Results

## 6.1. Geometry of vision and sub-Riemannian geometry: new results

Let us list here our new results in the geometry of vision axis and, more generally, on hypoelliptic diffusion and sub-Riemannian geometry.

- In [7] we present a new image inpainting algorithm, the Averaging and Hypoelliptic Evolution (AHE) algorithm, inspired by the one presented in [86] and based upon a (semi-discrete) variation of the Citti–Petitot–Sarti model of the primary visual cortex V1. In particular, we focus on reconstructing highly corrupted images (i.e. where more than the 80% of the image is missing).
- In [6] we deal with a severe ill posed problem, namely the reconstruction process of an image during tomography acquisition with (very) few views. We present different methods that we have been investigated during the past decade. They are based on variational analysis.

- [13] is the first paper of a series in which we plan to study spectral asymptotics for sub-Riemannian Laplacians and to extend results that are classical in the Riemannian case concerning Weyl measures, quantum limits, quantum ergodicity, quasi-modes, trace formulae. Even if hypoelliptic operators have been well studied from the point of view of PDEs, global geometrical and dynamical aspects have not been the subject of much attention. As we will see, already in the simplest case, the statements of the results in the sub-Riemannian setting are quite different from those in the Riemannian one. Let us consider a sub-Riemannian (sR) metric on a closed three-dimensional manifold with an oriented contact distribution. There exists a privileged choice of the contact form, with an associated Reeb vector field and a canonical volume form that coincides with the Popp measure. We establish a Quantum Ergodicity (QE) theorem for the eigenfunctions of any associated sR Laplacian under the assumption that the Reeb flow is ergodic. The limit measure is given by the normalized Popp measure.This is the first time that such a result is established for a hypoelliptic operator, whereas the usual Shnirelman theorem yields QE for the Laplace-Beltrami operator on a closed Riemannian manifold with ergodic geodesic flow. To prove our theorem, we first establish a microlocal Weyl law, which allows us to identify the limit measure and to prove the microlocal concentration of the eigenfunctions on the characteristic manifold of the sR Laplacian. Then, we derive a Birkhoff normal form along this characteristic manifold, thus showing that, in some sense, all 3D contact structures are microlocally equivalent. The quantum version of this normal form provides a useful microlocal factorization of the sR Laplacian. Using the normal form, the factorization and the ergodicity assumption, we finally establish a variance estimate, from which QE follows. We also obtain a second result, which is valid without any ergodicity assumption: every Quantum Limit (QL) can be decomposed in a sum of two mutually singular measures: the first measure is supported on the unit cotangent bundle and is invariant under the sR geodesic flow, and the second measure is supported on the characteristic manifold of the sR Laplacian and is invariant under the lift of the Reeb flow. Moreover, we prove that the first measure is zero for most QLs.
- In [22] we study the validity of the Whitney $C^1$ extension property for horizontal curves in sub-Riemannian manifolds endowed with 1-jets that satisfy a first-order Taylor expansion compatibility condition. We first consider the equiregular case, where we show that the extension property holds true whenever a suitable non-singularity property holds for the input-output maps on the Carnot groups obtained by nilpotent approximation. We then discuss the case of sub-Riemannian manifolds with singular points and we show that all step-2 manifolds satisfy the $C^1$ extension property. We conclude by showing that the $C^1$ extension property implies a Lusin-like approximation theorem for horizontal curves on sub-Riemannian manifolds.
- In [34] we prove the $C^1$ regularity for a class of abnormal length-minimizers in rank 2 sub-Riemannian structures. As a consequence of our result, all length-minimizers for rank 2 sub-Riemannian structures of step up to 4 are of class $C^1$.
- In [45] we address the double bubble problem for the anisotropic Grushin perimeter $P_\alpha$, $\alpha \geq 0$, and the Lebesgue measure in $\mathbb{R}^2$, in the case of two equal volumes. We assume that the contact interface between the bubbles lays on either the vertical or the horizontal axis. Since no regularity theory is available in this setting, in both cases we first prove existence of minimizers via the direct method by symmetrization arguments and then characterize them in terms of the given area by first variation techniques. Angles at which minimal boundaries intersect satisfy the standard 120-degree rule up to a suitable change of coordinates. While for $\alpha = 0$ the Grushin perimeter reduces to the Euclidean one and both minimizers coincide with the symmetric double bubble found in [104], for $\alpha = 1$ vertical interface minimizers have Grushin perimeter strictly greater than horizontal interface minimizers. As the latter ones are obtained by translating and dilating the Grushin isoperimetric set found in [131], we conjecture that they solve the double bubble problem with no assumptions on the contact interface.
- In [51] we study the notion of geodesic curvature of smooth horizontal curves parametrized by arc-length in the Heisenberg group, that is the simplest sub-Riemannian structure. Our goal is to give a metric interpretation of this notion of geodesic curvature as the first corrective term in the Taylor

expansion of the distance between two close points of the curve.

We would also like to mention the defense of the PhD thesis of Ludovic Saccchelli [3] on the subject.

## 6.2. Quantum control: new results

Let us list here our new results in quantum control theory.

- In [5] we consider a quantum particle in a potential $V(x)$ ($x \in \mathbb{R}^N$) in a time-dependent electric field $E(t)$ (the control). Boscain, Caponigro, Chambrion and Sigalotti proved in [83] that, under generic assumptions on $V$, this system is approximately controllable on the $L^2(\mathbb{R}^N, \mathbb{C})$-sphere, in sufficiently large time $T$. In the present article we show that approximate controllability does not hold in arbitrarily small time, no matter what the initial state is. This generalizes our previous result for Gaussian initial conditions. Moreover, we prove that the minimal time can in fact be arbitrarily large.
- In [11] we consider the bilinear Schrödinger equation with discrete-spectrum drift. We show, for $n \in \mathbb{N}$ arbitrary, exact controllability in projections on the first $n$ given eigenstates. The controllability result relies on a generic controllability hypothesis on some associated finite-dimensional approximations. The method is based on Lie-algebraic control techniques applied to the finite-dimensional approximations coupled with classical topological arguments issuing from degree theory.
- In [14] we consider the one dimensional Schrödinger equation with a bilinear control and prove the rapid stabilization of the linearized equation around the ground state. The feedback law ensuring the rapid stabilization is obtained using a transformation mapping the solution to the linearized equation on the solution to an exponentially stable target linear equation. A suitable condition is imposed on the transformation in order to cancel the non-local terms arising in the kernel system. This conditions also insures the uniqueness of the transformation. The continuity and invertibility of the transformation follows from exact controllability of the linearized system.
- In [33] we discuss how to control a parameter-dependent family of quantum systems. Our technique is based on adiabatic approximation theory and on the presence of curves of conical eigenvalue intersections of the controlled Hamiltonian. As particular cases, we recover chirped pulses for two-level quantum systems and counter-intuitive solutions for three-level stimulated Raman adiabatic passage (STIRAP). The proposed technique works for systems evolving both in finite-dimensional and infinite-dimensional Hilbert spaces. We show that the assumptions guaranteeing ensemble controllability are structurally stable with respect to perturbations of the parametrized family of systems.

## 6.3. Stability and uncertain dynamics: new results

Let us list here our new results about stability and stabilization of control systems, on the properties of systems with uncertain dynamics.

- In [8] we consider a one-dimensional controlled reaction-diffusion equation, where the control acts on the boundary and is subject to a constant delay. Such a model is a paradigm for more general parabolic systems coupled with a transport equation. We prove that it is possible to stabilize (in $H^1$ norm) this process by means of an explicit predictor-based feedback control that is designed from a finite-dimensional subsystem. The implementation is very simple and efficient and is based on standard tools of pole-shifting. Our feedback acts on the system as a finite-dimensional predictor. We compare our approach with the backstepping method.
- In [14] we consider the one dimensional Schrödinger equation with a bilinear control and prove the rapid stabilization of the linearized equation around the ground state. The feedback law ensuring the rapid stabilization is obtained using a transformation mapping the solution of the linearized equation to the solution of an exponentially stable target linear equation. A suitable condition is imposed on the transformation in order to cancel the non-local terms arising in the kernel system. This conditions also insures the uniqueness of the transformation. The continuity and invertibility of the transformation follows from exact controllability of the linearized system.

- Based on the notion of generalized homogeneity, we develop in [17] a new algorithm of feedback control design for a plant modeled by a linear evolution equation in a Hilbert space with a possibly unbounded operator. The designed control law steers any solution of the closed-loop system to zero in a finite time. Method of homogeneous extension is presented in order to make the developed control design principles to be applicable for evolution systems with non-homogeneous operators. The design scheme is demonstrated for heat equation with the control input distributed on the segment $[0, 1]$.
- In [19] we analyse the asymptotic behaviour of integro-differential equations modeling $N$ populations in interaction, all structured by different traits. Interactions are modeled by non-local terms involving linear combinations of the total number of individuals in each population. These models have already been shown to be suitable for the modeling of drug resistance in cancer, and they generalise the usual Lotka–Volterra ordinary differential equations. Our aim is to give conditions under which there is persistence of all species. Through the analysis of a Lyapunov function, our first main result gives a simple and general condition on the matrix of interactions, together with a convergence rate. The second main result establishes another type of condition in the specific case of mutualistic interactions. When either of these conditions is met, we describe which traits are asymptotically selected.
- The goal of [20] is to compute a boundary control of reaction-diffusion partial differential equation. The boundary control is subject to a constant delay, whereas the equation may be unstable without any control. For this system equivalent to a parabolic equation coupled with a transport equation, a prediction-based control is explicitly computed. To do that we decompose the infinite-dimensional system into two parts: one finite-dimensional unstable part, and one stable infinite-dimensional part. A finite-dimensional delay controller is computed for the unstable part, and it is shown that this controller succeeds in stabilizing the whole partial differential equation. The proof is based on a an explicit form of the classical Artstein transformation, and an appropriate Lyapunov function. A numerical simulation illustrate the constructive design method.
- [27] focuses on the (local) small-time stabilization of a Korteweg-de Vries equation on bounded interval, thanks to a time-varying Dirichlet feedback law on the left boundary. Recently, backstepping approach has been successfully used to prove the null controllability of the corresponding linearized system, instead of Carleman inequalities. We use the "adding an integrator" technique to gain regularity on boundary control term which clears the difficulty from getting stabilization in small-time.
- Motivated by improved ways to disrupt brain oscillations linked to Parkinson's disease, we propose in [29] an adaptive output feedback strategy for the stabilization of nonlinear time-delay systems evolving on a bounded set. To that aim, using the formalism of input-to-output stability (IOS), we first show that, for such systems, internal stability guarantees robustness to exogenous disturbances. We then use this feature to establish a general result on scalar adaptive output feedback of time-delay systems inspired by the "$\sigma$-modification" strategy. We finally apply this result to a delayed neuronal population model and assess numerically the performance of the adaptive stimulation.
- In [35] we consider open channels represented by Saint-Venant equations that are monitored and controlled at the downstream boundary and subject to unmeasured flow disturbances at the upstream boundary. We address the issue of feedback stabilization and disturbance rejection under Proportional-Integral (PI) boundary control. For channels with uniform steady states, the analysis has been carried out previously in the literature with spectral methods as well as with Lyapunov functions in Riemann coordinates. In [35], our main contribution is to show how the analysis can be extended to channels with non-uniform steady states with a Lyapunov function in physical coordinates.
- In [37], we study the exponential stabilization of a shock steady state for the inviscid Burgers equation on a bounded interval. Our analysis relies on the construction of an explicit strict control Lyapunov function. We prove that by appropriately choosing the feedback boundary conditions, we can stabilize the state as well as the shock location to the desired steady state in $H^2$-norm, with an arbitrary decay rate.

- Given a discrete-time linear switched system $\Sigma(A)$ associated with a finite set $A$ of matrices, we consider in [40] the measures of its asymptotic behavior given by, on the one hand, its deterministic joint spectral radius $\rho_d(A)$ and, on the other hand, its probabilistic joint spectral radii $\rho_p(v, P, A)$ for Markov random switching signals with transition matrix $P$ and a corresponding invariant probability $v$. Note that $\rho_d(A)$ is larger than or equal to $\rho_p(v, P, A)$ for every pair $(v, P)$. In this paper, we investigate the cases of equality of $\rho_d(A)$ with either a single $\rho_p(v, P, A)$ or with the supremum of $\rho_p(v, P, A)$ over $(v, P)$ and we aim at characterizing the sets $A$ for which such equalities may occur.
- In [41], we introduce a method to get necessary and sufficient stability conditions for systems governed by 1-D nonlinear hyperbolic partial-differential equations with closed-loop integral controllers, when the linear frequency analysis cannot be used anymore. We study the stability of a general nonlinear transport equation where the control input and the measured output are both located on the boundaries. The principle of the method is to extract the limiting part of the stability from the solution using a projector on a finite-dimensional space and then use a Lyapunov approach. We improve a result of Trinh, Andrieu and Xu, and give an optimal condition for the design of the controller. The results are illustrated with numerical simulations where the predicted stable and unstable regions can be clearly identified.
- In [44] we construct explicit time-varying feedback laws leading to the global (null) stabilization in small time of the viscous Burgers equation with three scalar controls. Our feedback laws use first the quadratic transport term to achieve the small-time global approximate stabilization and then the linear viscous term to get the small-time local stabilization.
- In [46] we address the question of the exponential stability for the $C^1$ norm of general 1-D quasilinear systems with source terms under boundary conditions. To reach this aim, we introduce the notion of basic $C^1$ Lyapunov functions, a generic kind of exponentially decreasing function whose existence ensures the exponential stability of the system for the $C^1$ norm. We show that the existence of a basic $C^1$ Lyapunov function is subject to two conditions: an interior condition, intrinsic to the system, and a condition on the boundary controls. We give explicit sufficient interior and boundary conditions such that the system is exponentially stable for the $C^1$ norm and we show that the interior condition is also necessary to the existence of a basic $C^1$ Lyapunov function. Finally, we show that the results conducted in this article are also true under the same conditions for the exponential stability in the $C^p$ norm, for any $p \geq 1$.
- In [47] we study the exponential stability for the $C^1$ norm of general $2\times2$ 1-D quasilinear hyperbolic systems with source terms and boundary controls. When the eigenvalues of the system have the same sign, any nonuniform steady-state can be stabilized using boundary feedbacks that only depend on measurements at the boundaries and we give explicit conditions on the gain of the feedback. In other cases, we exhibit a simple numerical criterion for the existence of basic $C^1$ Lyapunov function, a natural candidate for a Lyapunov function to ensure exponential stability for the $C^1$ norm. We show that, under a simple condition on the source term, the existence of a basic $C^1$ (or $C^p$ , for any $p \geq 1$) Lyapunov function is equivalent to the existence of a basic $H^2$ (or $H^q$ , for any $q \geq 2$) Lyapunov function, its analogue for the $H^2$ norm. Finally, we apply these results to the nonlinear Saint-Venant equations. We show in particular that in the subcritical regime, when the slope is larger than the friction, the system can always be stabilized in the $C^1$ norm using static boundary feedbacks depending only on measurements of at the boundaries, which has a large practical interest in hydraulic and engineering applications.
- In [48] we study the exponential stability in the $H^2$ norm of the nonlinear Saint-Venant (or shallow water) equations with arbitrary friction and slope using a single Proportional-Integral (PI) control at one end of the channel. Using a local dissipative entropy we find a simple and explicit condition on the gain the PI control to ensure the exponential stability of any steady-states. This condition is independent of the slope, the friction, the length of the river, the inflow disturbance and, more surprisingly, the steady-state considered. When the inflow disturbance is time-dependent and no steady-state exist, we still have the Input-to-State stability of the system, and we show that changing

slightly the PI control enables to recover the exponential stability of slowly varying trajectories.

- The exponential stability problem of the nonlinear Saint-Venant equations is addressed in [49]. We consider the general case where an arbitrary friction and space-varying slope are both included in the system, which lead to non-uniform steady-states. An explicit quadratic Lyapunov function as a weighted function of a small perturbation of the steady-states is constructed. Then we show that by a suitable choice of boundary feedback controls, that we give explicitly, the local exponential stability of the nonlinear Saint-Venant equations for the $H^2$-norm is guaranteed.
- [53] elaborates control strategies to prevent clustering effects in opinion formation models. This is the exact opposite of numerous situations encountered in the literature where, on the contrary, one seeks controls promoting consensus. In order to promote declustering, instead of using the classical variance that does not capture well the phenomenon of dispersion, we introduce an entropy-type functional that is adapted to measuring pairwise distances between agents. We then focus on a Hegselmann-Krause-type system and design declustering sparse controls both in finite-dimensional and kinetic models. We provide general conditions characterizing whether clustering can be avoided as function of the initial data. Such results include the description of black holes (where complete collapse to consensus is not avoidable), safety zones (where the control can keep the system far from clustering), basins of attraction (attractive zones around the clustering set) and collapse prevention (when convergence to the clustering set can be avoided).
- In [54] we consider the problem of controlling parabolic semilinear equations arising in population dynamics, either in finite time or infinite time. These are the monostable and bistable equations on $(0, L)$ for a density of individuals $0 \leq y(t, x) \leq 1$, with Dirichlet controls taking their values in $[0, 1]$. We prove that the system can never be steered to extinction (steady state 0) or invasion (steady state 1) in finite time, but is asymptotically controllable to 1 independently of the size $L$, and to 0 if the length $L$ of the interval domain is less than some threshold value $L^*$, which can be computed from transcendental integrals. In the bistable case, controlling to the other homogeneous steady state $0 < \theta < 1$ is much more intricate. We rely on a staircase control strategy to prove that $\theta$ can be reached in finite time if and only if $L < L^?$. The phase plane analysis of those equations is instrumental in the whole process. It allows us to read obstacles to controllability, compute the threshold value for domain size as well as design the path of steady states for the control strategy.
- Given a linear control system in a Hilbert space with a bounded control operator, we establish in [56] a characterization of exponential stabilizability in terms of an observability inequality. Such dual characterizations are well known for exact (null) controllability. Our approach exploits classical Fenchel duality arguments and, in turn, leads to characterizations in terms of observability inequalities of approximately null controllability and of $\alpha$-null controllability. We comment on the relationships between those various concepts, at the light of the observability inequalities that characterize them.
- In [58] we use the backstepping method to study the stabilization of a 1-D linear transport equation on the interval $(0, L)$, by controlling the scalar amplitude of a piecewise regular function of the space variable in the source term. We prove that if the system is controllable in a periodic Sobolev space of order greater than 1, then the system can be stabilized exponentially in that space and, for any given decay rate, we give an explicit feedback law that achieves that decay rate.

Let us also mention the lecture notes [31] on stabilization of semilinear PDE's, which have been published this year.

## 6.4. Optimal control: new results

Let us list here our new results in optimal control theory beyond the sub-Riemannian framework.

- In [4] we focus on regional deterministic optimal control problems, i.e., problems where the dynamics and the cost functional may be different in several regions of the state space and present discontinuities at their interface. Under the assumption that optimal trajectories have a locally finite

number of switchings (no Zeno phenomenon), we use the duplication technique to show that the value function of the regional optimal control problem is the minimum over all possible structures of trajectories of value functions associated with classical optimal control problems settled over fixed structures, each of them being the restriction to some submanifold of the value function of a classical optimal control problem in higher dimension. The lifting duplication technique is thus seen as a kind of desingularization of the value function of the regional optimal control problem. In turn, we extend to regional optimal control problems the classical sensitivity relations and we prove that the regularity of this value function is the same (i.e., is not more degenerate) than the one of the higher-dimensional classical optimal control problem that lifts the problem.

- The goal of [9] is to show how non-parametric statistics can be used to solve some chance constrained optimization and optimal control problems. We use the Kernel Density Estimation method to approximate the probability density function of a random variable with unknown distribution, from a relatively small sample. We then show how this technique can be applied and implemented for a class of problems including the Goddard problem and the trajectory optimization of an Ariane 5-like launcher.
- In control theory the term chattering is used to refer to fast oscillations of controls, such as an infinite number of switchings over a finite time interval. In [10] we focus on three typical instances of chattering: the Fuller phenomenon, referring to situations where an optimal control features an accumulation of switchings in finite time; the Robbins phenomenon, concerning optimal control problems with state constraints, where the optimal trajectory touches the boundary of the constraint set an infinite number of times over a finite time interval; and the Zeno phenomenon, for hybrid systems, referring to a trajectory that depicts an infinite number of location switchings in finite time. From the practical point of view, when trying to compute an optimal trajectory, for instance, by means of a shooting method, chattering may be a serious obstacle to convergence. In [10] we propose a general regularization procedure, by adding an appropriate penalization of the total variation. This produces a family of quasi-optimal controls whose associated cost converge to the optimal cost of the initial problem as the penalization tends to zero. Under additional assumptions, we also quantify quasi-optimality by determining a speed of convergence of the costs.
- In [12], a new robust and fast method is developed to perform transfers that minimize fuel consumption between two invariant manifolds of periodic orbits in the circular restricted three-body problem. The method starts with an impulse transfer between two invariant manifolds to build an optimal control problem. This allows to choose an adequate fixed transfer time. Using the Pontryagin maximum principle, the resolution of the problem is formulated as that of finding the zero of a shooting function (indirect method). The algorithm couples different kinds of continuations (on cost, final state, and thrust) to improve robustness and to initialize the solver. The efficiency of the method is illustrated with numerical examples. Finally, the influence of the transfer time is studied numerically thanks to a continuation on this parameter, and it checks that, when transfer duration goes to zero, the control converges to the impulse transfer that it started with. It shows the robustness of the method and establishes a mathematical link between the two problems.
- In [15] we consider the controllability problem for finite-dimensional linear autonomous control systems, under state constraints but without imposing any control constraint. It is well known that, under the classical Kalman condition, in the absence of constraints on the state and the control, one can drive the system from any initial state to any final one in an arbitrarily small time. Furthermore, it is also well known that there is a positive minimal time in the presence of compact control constraints. We prove that, surprisingly, a positive minimal time may be required as well under state constraints, even if one does not impose any restriction on the control. This may even occur when the state constraints are unilateral, like the nonnegativity of some components of the state, for instance. Using the Brunovsky normal forms of controllable systems, we analyze this phenomenon in detail, that we illustrate by several examples. We discuss some extensions to nonlinear control systems and formulate some challenging open problems.

- In [18] we consider a system of two coupled integro-differential equations modeling populations of healthy and cancer cells under therapy. Both populations are structured by a phenotypic variable, representing their level of resistance to the treatment. We analyse the asymptotic behaviour of the model under constant infusion of drugs. By designing an appropriate Lyapunov function, we prove that both densities converge to Dirac masses. We then define an optimal control problem, by considering all possible infusion protocols and minimising the number of cancer cells over a prescribed time frame. We provide a quasi-optimal strategy and prove that it solves this problem for large final times. For this modeling framework, we illustrate our results with numerical simulations, and compare our optimal strategy with periodic treatment schedules.
- In [21] we use conductance based neuron models and the mathematical modeling of Optogenetics to define controlled neuron models and we address the minimal time control of these affine systems for the first spike from equilibrium. We apply tools of geometric optimal control theory to study singular extremals and we implement a direct method to compute optimal controls. When the system is too large to theoretically investigate the existence of singular optimal controls, we observe numerically the optimal bang-bang controls.
- In [23] we first derive a general integral-turnpike property around a set for infinite-dimensional non-autonomous optimal control problems with any possible terminal state constraints, under some appropriate assumptions. Roughly speaking, the integral-turnpike property means that the time average of the distance from any optimal trajectory to the turnpike set converges to zero, as the time horizon tends to infinity. Then, we establish the measure-turnpike property for strictly dissipative optimal control systems, with state and control constraints. The measure-turnpike property, which is slightly stronger than the integral-turnpike property, means that any optimal (state and control) solution remains essentially, along the time frame, close to an optimal solution of an associated static optimal control problem, except along a subset of times that is of small relative Lebesgue measure as the time horizon is large. Next, we prove that strict strong duality, which is a classical notion in optimization, implies strict dissipativity, and measure-turnpike. Finally, we conclude the paper with several comments and open problems.
- In [24], we investigate the asymptotic behavior of optimal designs for the shape optimization of 2D heat equations in long time horizons. The control is the shape of the domain on which heat diffuses. The class of 2D admissible shapes is the one introduced by Sverák, of all open subsets of a given bounded open set, whose complementary sets have a uniformly bounded number of connected components. Using a $\Gamma$-convergence approach, we establish that the parabolic optimal designs converge as the length of the time horizon tends to infinity, in the complementary Hausdorff topology, to an optimal design for the corresponding stationary elliptic equation.
- In [25], we study the steady-state (or periodic) exponential turnpike property of optimal control problems in Hilbert spaces. The turnpike property, which is essentially due to the hyperbolic feature of the Hamiltonian system resulting from the Pontryagin maximum principle, reflects the fact that, in large time, the optimal state, control and adjoint vector remain most of the time close to an optimal steady-state. A similar statement holds true as well when replacing an optimal steady-state by an optimal periodic trajectory. To establish the result, we design an appropriate dichotomy transformation, based on solutions of the algebraic Riccati and Lyapunov equations. We illustrate our results with examples including linear heat and wave equations with periodic tracking terms.
- The Allee threshold of an ecological system distinguishes the sign of population growth either towards extinction or to carrying capacity. In practice human interventions can tune the Allee threshold for instance thanks to the sterile male technique and the mating disruption. In [26] we address various control objectives for a system described by a diffusion-reaction equation regulating the Allee threshold, viewed as a real parameter determining the unstable equilibrium of the bistable nonlinear reaction term. We prove that this system is the mean field limit of an interacting system of particles in which individual behaviours are driven by stochastic laws. Numerical simulations of the stochastic process show that population propagations are governed by wave-like solutions corresponding to traveling solutions of the macroscopic reaction-diffusion system. An optimal

control problem for the macroscopic model is then introduced with the objective of steering the system to a target traveling wave. The relevance of this problem is motivated by the fact that traveling wave solutions model the fact that bounded space domains reach asymptotically an equilibrium configuration. Using well known analytical results and stability properties of traveling waves, we show that well-chosen piecewise constant controls allow to reach the target approximately in sufficiently long time. We then develop a direct computational method and show its efficiency for computing such controls in various numerical simulations. Finally we show the efficiency of the obtained macroscopic optimal controls in the microscopic system of interacting particles and we discuss their advantage when addressing situations that are out of reach for the analytical methods. We conclude the article with some open problems and directions for future research.

- Consider a general nonlinear optimal control problem in finite dimension, with constant state and/or control delays. By the Pontryagin Maximum Principle, any optimal trajectory is the projection of a Pontryagin extremal. In [39] we establish that, under appropriate assumptions, Pontryagin extremals depend continuously on the parameter delays, for adequate topologies. The proof of the continuity of the trajectory and of the control is quite easy, however, for the adjoint vector, the proof requires a much finer analysis. The continuity property of the adjoint with respect to the parameter delay opens a new perspective for the numerical implementation of indirect methods, such as the shooting method. We also discuss the sharpness of our assumptions.
- In [43] we are concerned about the controllability of a general linear hyperbolic system of the form $\partial_t w(t,x) = \Sigma(x)\partial_x w(t,x) + \gamma C(x) w(t,x)$ ($\gamma \in \mathbb{R}$) in one space dimension using boundary controls on one side. More precisely, we establish the optimal time for the null and exact controllability of the hyperbolic system for generic $\gamma$. We also present examples which yield that the generic requirement is necessary. In the case of constant $\Sigma$ and of two positive directions, we prove that the null-controllability is attained for any time greater than the optimal time for all $\gamma \in \mathbb{R}$ and for all $C$ which is analytic if the slowest negative direction can be alerted by both positive directions. We also show that the null-controllability is attained at the optimal time by a feedback law when $C \equiv 0$. Our approach is based on the backstepping method paying a special attention on the construction of the kernel and the selection of controls.
- In [52] we consider a state-constrained optimal control problem of a system of two non-local partial-differential equations, which is an extension of the one introduced in a previous work in mathematical oncology. The aim is to minimize the tumor size through chemotherapy while avoiding the emergence of resistance to the drugs. The numerical approach to solve the problem was the combination of direct methods and continuation on discretization parameters, which happen to be insufficient for the more complicated model, where diffusion is added to account for mutations. In [52], we propose an approach relying on changing the problem so that it can theoretically be solved thanks to a Pontryagin Maximum Principle in infinite dimension. This provides an excellent starting point for a much more reliable and efficient algorithm combining direct methods and continuations. The global idea is new and can be thought of as an alternative to other numerical optimal control techniques.

We would also like to mention the defense of the PhD theses of Riccardo Bonalli [1] and Antoine Olivier [2] on the subject.

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Bilateral Contracts with Industry

A bilateral contract with CNES funded the PhD thesis of Antoine Olivier, who defended in October 2018.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

### *8.1.1. ANR*

- ANR SRGI, for *Sub-Riemannian Geometry and Interactions*, coordinated by **Emmanuel Trélat**, started in 2015 and runs until 2020. Other partners: Toulon University and Grenoble University. SRGI deals with sub-Riemannian geometry, hypoelliptic diffusion and geometric control.
- ANR Finite4SoS, for *Commande et estimation en temps fini pour les Systèmes de Systèmes*, coordinated by Wilfrid Perruquetti, started in 2015 and runs until 2019. Other partners: Inria Lille, CAOR - ARMINES. Finite4SoS aims at developing a new promising framework to address control and estimation issues of Systems of Systems subject to model diversity, while achieving robustness as well as severe time response constraints.
- ANR QUACO, for *QUAntum COntrol: PDE systems and MRI applications*, coordinated by Thomas Chambrion, started in 2017 and runs until 2021. Other partners: Lorraine University. QUACO aims at contributing to quantum control theory in two directions: improving the comprehension of the dynamical properties of controlled quantum systems in infinite-dimensional state spaces, and improve the efficiency of control algorithms for MRI.

## 8.2. European Initiatives

### *8.2.1. H2020 Projects*

Program: ERC Proof of Concept

Project acronym: ARTIV1

Project title: An artificial visual cortex for image processing

Duration: From April 2017 to September 2018.

Coordinator: Ugo Boscain

Abstract: The ERC starting grant GECOMETHODS, on which this POC is based, tackled problems of diffusion equations via geometric control methods. One of the most striking achievements of the project has been the development of an algorithm of image reconstruction based mainly on non-isotropic diffusion. This algorithm is bio-mimetic in the sense that it replicates the way in which the primary visual cortex V1 of mammals processes the signals arriving from the eyes. It has performances that are at the state of the art in image processing. These results together with others obtained in the ERC project show that image processing algorithms based on the functional architecture of V1 can go very far. However, the exceptional performances of the primary visual cortex V1 rely not only on the particular algorithm used, but also on the fact that such algorithm 'runs' on a dedicated hardware having the following features: 1. an exceptional level of parallelism; 2. connections that are well adapted to transmit information in a non-isotropic way as it is required by the algorithms of image reconstruction and recognition. The idea of this POC is to create a dedicated hardware (called ARTIV1) emulating the functional architecture of V1 and hence having on one hand a huge degree of parallelism and on the other hand connections among the CPUs that reflect the non-isotropic structure of the visual cortex V1.

## 8.3. International Research Visitors

### *8.3.1. Research Stays Abroad*

Jean-Michel Coron was at EPFL (Switzerland) from January to June 2018.

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Organisation*

*9.1.1.1. Member of the Organizing Committees*

Ugo Boscain and Mario Sigalotti were Members of the Organizing Committee of the Workshop “Sub-Riemannian Geometry and Topolò(gy)", Topolò/Topolove, Italy, June 2018

### *9.1.2. Scientific Events Selection*

*9.1.2.1. Member of the Conference Program Committees*

- Emmanuel Trélat was Member of the Program Committee of the 18th French-German-Italian Conference on Optimization (FGI’2018).
- Emmanuel Trélat was Member of the Scientific Committee of the 23rd International Symposium on Mathematical Programming (ISMP 2018).

### *9.1.3. Journal*

*9.1.3.1. Member of the Editorial Boards*

- Ugo Boscain is Associate editor of SIAM Journal of Control and Optimization
- Ugo Boscain is Managing editor of Journal of Dynamical and Control Systems
- Jean-Michel Coron is Editor-in-chief of Comptes Rendus Mathématique
- Jean-Michel Coron is Member of the editorial board of Journal of Evolution Equations
- Jean-Michel Coron is Member of the editorial board of Asymptotic Analysis
- Jean-Michel Coron is Member of the editorial board of ESAIM : Control, Optimisation and Calculus of Variations
- Jean-Michel Coron is Member of the editorial board of Applied Mathematics Research Express
- Jean-Michel Coron is Member of the editorial board of Advances in Differential Equations
- Jean-Michel Coron is Member of the editorial board of Math. Control Signals Systems
- Jean-Michel Coron is Member of the editorial board of Annales de l’IHP, Analyse non linéaire
- Mario Sigalotti is Associate editor of ESAIM : Control, Optimisation and Calculus of Variations
- Mario Sigalotti is Associate editor of Journal on Dynamical and Control Systems
- Emmanuel Trélat is Editor-in-chief of ESAIM : Control, Optimisation and Calculus of Variations
- Emmanuel Trélat is Associate editor of Syst. Cont. Letters
- Emmanuel Trélat is Associate editor of J. Dynam. Cont. Syst.
- Emmanuel Trélat is Associate editor of Bollettino dell’Unione Matematica Italiana
- Emmanuel Trélat is Associate editor of ESAIM Math. Modelling Num. Analysis
- Emmanuel Trélat is Editor of BCAM Springer Briefs
- Emmanuel Trélat is Associate editor of J. Optim. Theory Appl.
- Emmanuel Trélat is Associate editor of Math. Control Related fields

### *9.1.4. Invited Talks*

- Ugo Boscain was invited speaker at the International Conference “Optimal Control and Differential Games”, dedicated to the 110th anniversary of L.S. Pontryagin, Dec. 2018.
- Ugo Boscain was invited speaker at the conference “Dynamics, Control, and Geometry", Banach Center, Warsaw, Sept. 2018.

- Ugo Boscain was invited speaker at Linkopyng University, Department of Electrical Engineering, Nov. 2018.
- Ugo Boscain was invited speaker at the conference “Analysis, Control and Inverse Problems for PDEs”, Napoli (Italy), Nov. 2018.
- Mario Sigalotti was invited speaker at the Workshop Quantum control and feedback: foundations and applications, Paris, Jun. 2018.
- Emmanuel Trélat was invited speaker at ICM 2018, Rio, section “Control Theory and Optimization", Aug. 2018.
- Emmanuel Trélat was invited speaker at Analysis, Control and Inverse Problems for PDEs, Naples, Nov. 2018.
- Emmanuel Trélat was invited speaker at Dynamics Control and Geometry, Varsovie, Sept. 2018.
- Emmanuel Trélat was invited speaker at 14th Viennese Conference on Optimal Control and Dynamic Games, Vienna, July 2018.
- Emmanuel Trélat was invited speaker at Portuguese Meeting on Optimal Control 2018, Coimbra (Portugal), June 2018.
- Emmanuel Trélat was invited speaker at International Symposium on Mathematical Control Theory, Shanghai, June 2018.
- Emmanuel Trélat was invited speaker at GAMM Munich, March 2018.

### *9.1.5. Leadership within the Scientific Community*

Emmanuel Trélat is director of the Fondation Sciences Mathématiques de Paris (FSMP).

## 9.2. Teaching - Supervision - Juries

### *9.2.1. Teaching*

- Ugo Boscain thought “Sub-elliptic diffusion” to PhD students at SISSA, Trieste Italy
- Ugo Boscain thought “Automatic Control” (with Mazyar Mirrahimi) at Ecole Polytechnique
- Ugo Boscain thought “MODAL of applied mathematics. Contrôle de modèles dynamiques” at Ecole Polytechnique
- Emmanuel Trélat thought “Control in finite and infinite dimension” at Master 2, Sorbonne Université

### *9.2.2. Supervision*

PhD: Riccardo Bonalli, Optimal Control of Aerospace Systems with Control-State Constraints and Delays, Sorbonne Université, July 2018, supervised by Emmanuel Trélat.

PhD: Antoine Olivier, Optimal and robust attitude control of a launcher, Sorbonne Université, October 2018, supervised by Emmanuel Trélat and co-supervised by Thomas Haberkorn, Éric Bourgeois, David-Alexis Handschuh.

PhD: Ludovic Sacchelli, Singularities in sub-Riemannian geometry, Université Paris-Saclay, September 2018, supervised by Ugo Boscain and Mario Sigalotti.

PhD in progress: Nicolas Augier, “Contrôle adiabatique des systèmes quantiques", started in September 2016, supervisors: Ugo Boscain, Mario Sigalotti.

PhD in progress: Amaury Hayat, “ Contrôle et stabilisation en mécanique des fluides", started in October 2016, supervisors: Jean-Michel Coron and Sébastien Boyaval

PhD in progress: Mathieu Kohli, “Volume and curvature in sub-Riemannian geometry", started in September 2016, supervisors: Davide Barilari, Ugo Boscain.

PhD in progress: Gontran Lance, started in September 2018, supervisors: Emmanuel Trélat and Enrique Zuazua.

PhD in progress: Cyril Letrouit, “Équation des ondes sous-riemanniennes", started in September 2019, supervisor Emmanuel Trélat.
PhD in progress: Jakub Orłowski, “Modeling and steering brain oscillations based on in vivo optogenetics data", started in September 2016, supervisors: Antoine Chaillet, Alain Destexhe, and Mario Sigalotti.
PhD in progress: Eugenio Pozzoli, “Adiabatic Control of Open Quantum Systems", started in September 2018, supervisors: Ugo Boscain and Mario Sigalotti.
PhD in progress: Shengquan Xiang, Stabilisation des fluides par feedbacks non-linéaires, September 2016, supervisor: Jean-Michel Coron.
PhD in progress: Christophe Zhang, started in October 2016, supervisor: Jean-Michel Coron

#### *9.2.3. Juries*

- Ugo Boscain was referee and member of the jury of the HDR of Jean-Marie Mirebeau, Université Paris-Sud.
- Mario Sigalotti was member of the jury of the PhD thesis of Abdelkrim Bahloul, Univ. Paris-Saclay.
- Emmanuel Trélat was co-supervisor and member of the jury of the PhD thesis of Camille Pouchol, Sorbonne Université.
- Emmanuel Trélat was member of the jury of the PhD thesis of F. Omnès, Sorbonne Université.
- Emmanuel Trélat was referee and member of the jury of the PhD thesis of S. Mitra, Univ. Toulouse.
- Emmanuel Trélat was referee and member of the jury of the PhD thesis of T. Weisser, Univ. Toulouse.
- Emmanuel Trélat was referee and member of the jury of the PhD thesis of S. Maslovskaya, Univ. Paris-Saclay.
- Emmanuel Trélat was referee and member of the jury of the PhD thesis of A. Vieira, Grenoble University.
- Emmanuel Trélat was member of the jury of the HDR of F. Chittaro, Univ. Toulon.

### 9.3. Popularization

Emmanuel Trélat is member of the Comité d’Honneur du Salon des Jeux et Culture Mathématique since November 2018

#### *9.3.1. Articles and contents*

- Nicolas Augier, Ugo Boscain, and Mario Sigalotti are authors of the popularization article [32] explaining how broken adiabatic paths can be used to enhance the control of a quantum systems.

#### *9.3.2. Education*

- Ugo Boscain and Jean-Michel Coron gave a lecture at journée ENS-UPS, ENS Paris
- Emmanuel Trélat gave a lecture at ENS Ulm to first-year students
- Emmanuel Trélat gave a lecture at Université Paris-Diderot to first- and second-year students

#### *9.3.3. Interventions*

- Ugo Boscain gave a lecture at Alfaclass, Saint-Barthélemy, Aosta, Italy
- Emmanuel Trélat gave a lecture at Salon des Jeux et Culture Mathématique

## 10. Bibliography

### Publications of the year

#### Doctoral Dissertations and Habilitation Theses

[1] R. BONALLI.*Optimal Control of Aerospace Systems with Control-State Constraints and Delays*, Sorbonne Université, UPMC University of Paris 6, Laboratoire Jacques-Louis Lions ; ONERA – The French Aerospace Lab, Département TIS, Unité NGPA ; Inria Paris, Equipe CAGE, July 2018, https://tel.archives-ouvertes.fr/tel-01848542

[2] A. OLIVIER.*Optimal and robust attitude control of a launcher*, Sorbonne Université, October 2018, https://tel.archives-ouvertes.fr/tel-01962542

[3] L. SACCHELLI.*Singularities in sub-Riemannian geometry*, Université Paris-Saclay, September 2018, https://pastel.archives-ouvertes.fr/tel-01893068

### Articles in International Peer-Reviewed Journal

[4] G. BARLES, A. BRIANI, E. TRÉLAT.*Value function for regional control problems via dynamic programming and Pontryagin maximum principle*, in "Mathematical Control and Related Fields", 2018, vol. 8, n$^{o}$ 3-4, p. 509–533, https://arxiv.org/abs/1605.04079 , https://hal.archives-ouvertes.fr/hal-01313559

[5] K. BEAUCHARD, J.-M. CORON, H. TEISMANN.*Minimal time for the approximate bilinear control of Schrödinger equations*, in "Mathematical Methods in the Applied Sciences", 2018, vol. 41, n$^{o}$ 5, p. 1831-1844 [*DOI :* 10.1002/MMA.4710], https://hal.archives-ouvertes.fr/hal-01333537

[6] M. BERGOUNIOUX, I. ABRAHAM, R. ABRAHAM, G. CARLIER, E. LE PENNEC, E. TRÉLAT.*Variational methods for tomographic reconstruction with few views*, in "Milan Journal of Mathematics", 2018, vol. 86, n$^{o}$ 2, p. 157–200, https://hal.archives-ouvertes.fr/hal-01817172

[7] U. BOSCAIN, R. CHERTOVSKIH, J.-P. GAUTHIER, D. PRANDI, A. REMIZOV.*Highly corrupted image inpainting through hypoelliptic diffusion*, in "Journal of Mathematical Imaging and Vision", April 2018, p. 1–15, https://arxiv.org/abs/1502.07331 [*DOI :* 10.1007/s10851-018-0810-4], https://hal.inria.fr/hal-01139521

[8] D. BRESCH-PIETRI, C. PRIEUR, E. TRÉLAT.*Finite-dimensional predictor-based feedback stabilization of a 1D linear reaction-diffusion equation with boundary input delay*, in "Systems and Control Letters", 2018, vol. 113, p. 9–16, https://arxiv.org/abs/1511.03030 , https://hal.archives-ouvertes.fr/hal-01226598

[9] J.-B. CAILLAU, M. CERF, A. SASSI, E. TRÉLAT, H. ZIDANI.*Solving chance constrained optimal control problems in aerospace via Kernel Density Estimation*, in "Optimal Control Appl. Methods", 2018, vol. 39, n$^{o}$ 5, p. 1833-1858 [*DOI :* 10.1002/OCA.2445], https://hal.inria.fr/hal-01507063

[10] M. CAPONIGRO, R. GHEZZI, B. PICCOLI, E. TRÉLAT.*Regularization of chattering phenomena via bounded variation controls*, in "IEEE Transactions on Automatic Control", July 2018, vol. 63, n$^{o}$ 7, p. 2046-2060, https://arxiv.org/abs/1303.5796 [*DOI :* 10.1109/TAC.2018.2810540], https://hal.archives-ouvertes.fr/hal-01359037

[11] M. CAPONIGRO, M. SIGALOTTI.*Exact controllability in projections of the bilinear Schrödinger equation*, in "SIAM Journal on Control and Optimization", 2018, vol. 56, https://hal.inria.fr/hal-01509971

[12] M. CHUPIN, T. HABERKORN, E. TRÉLAT.*Transfer Between Invariant Manifolds: From Impulse Transfer to Low-Thrust Transfer*, in "Journal of Guidance, Control, and Dynamics", 2018, vol. 41, n$^{o}$ 3, p. 658–672 [*DOI :* 10.2514/1.G002922], https://hal.inria.fr/hal-01494042

[13] Y. COLIN DE VERDIÈRE, L. HILLAIRET, E. TRÉLAT.*Spectral asymptotics for sub-Riemannian Laplacians. I: Quantum ergodicity and quantum limits in the 3D contact case*, in "Duke Mathematical Journal", 2018, vol. 167, n$^{o}$ 1, p. 109–174, https://arxiv.org/abs/1504.07112 , https://hal.archives-ouvertes.fr/hal-01144257

[14] J.-M. CORON, L. GAGNON, M. MORANCEY.*Rapid Stabilization of a Linearized Bilinear 1-D Schrödinger Equation*, in "Journal de Mathématiques Pures et Appliquées", July 2018, vol. 115, https://hal.archives-ouvertes.fr/hal-01408179

[15] J. LOHÉAC, E. TRÉLAT, E. ZUAZUA.*Minimal controllability time for finite-dimensional control systems under state constraints*, in "Automatica", October 2018, vol. 96, p. 380-392 [*DOI :* 10.1016/J.AUTOMATICA.2018.07.010], https://hal.archives-ouvertes.fr/hal-01710759

[16] A. OLIVIER, T. HABERKORN, E. TRÉLAT, E. BOURGEOIS, D.-A. HANDSCHUH.*Redundancy implies robustness for bang-bang strategies*, in "Optimal Control Applications and Methods", 2019, vol. 40, n$^{o}$ 1, p. 85–104, https://arxiv.org/abs/1707.02053 , https://hal.archives-ouvertes.fr/hal-01557937

[17] 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

[18] C. POUCHOL, J. CLAIRAMBAULT, A. LORZ, E. TRÉLAT.*Asymptotic analysis and optimal control of an integro-differential system modelling healthy and cancer cells exposed to chemotherapy*, in "Journal de Mathématiques Pures et Appliquées", 2018, vol. 116, p. 268–308, https://arxiv.org/abs/1612.04698 [*DOI :* 10.1016/J.MATPUR.2017.10.007], https://hal.archives-ouvertes.fr/hal-01673589

[19] C. POUCHOL, E. TRÉLAT.*Global stability with selection in integro-differential Lotka-Volterra systems modelling trait-structured populations*, in "Journal of Biological Dynamics", 2018, vol. 12, n$^{o}$ 1, p. 872–893, https://arxiv.org/abs/1702.06187 , https://hal.archives-ouvertes.fr/hal-01470722

[20] C. PRIEUR, E. TRÉLAT.*Feedback stabilization of a 1D linear reaction-diffusion equation with delay boundary control*, in "IEEE Transactions on Automatic Control", 2018, https://arxiv.org/abs/1709.02735 [*DOI :* 10.1109/TAC.2018.2849560], https://hal.archives-ouvertes.fr/hal-01583199

[21] V. RENAULT, M. THIEULLEN, E. TRÉLAT.*Minimal time spiking in various ChR2-controlled neuron models*, in "Journal of Mathematical Biology", 2018, vol. 76, n$^{o}$ 3, p. 567–608, https://hal.archives-ouvertes.fr/hal-01320492

[22] L. SACCHELLI, M. SIGALOTTI.*On the Whitney extension property for continuously differentiable horizontal curves in sub-Riemannian manifolds*, in "Calculus of Variations and Partial Differential Equations", 2018, https://arxiv.org/abs/1708.02795 , https://hal.archives-ouvertes.fr/hal-01573353

[23] E. TRÉLAT, C. ZHANG.*Integral and measure-turnpike properties for infinite-dimensional optimal control systems*, in "Mathematics of Control, Signals, and Systems", 2018, vol. 30, n$^{o}$ 1, 30:3, https://arxiv.org/abs/1705.02762 , https://hal.archives-ouvertes.fr/hal-01519490

[24] E. TRÉLAT, C. ZHANG, E. ZUAZUA.*Optimal shape design for 2D heat equations in large time*, in "Pure and Applied Functional Analysis", 2018, vol. 3, n$^{o}$ 1, p. 255–269, https://hal.archives-ouvertes.fr/hal-01442997

[25] E. TRÉLAT, C. ZHANG, E. ZUAZUA.*Steady-state and periodic exponential turnpike property for optimal control problems in Hilbert spaces*, in "SIAM Journal on Control and Optimization", 2018, vol. 56, n$^{o}$ 2, p. 1222-1252, https://hal.archives-ouvertes.fr/hal-01377320

[26] E. TRÉLAT, J. ZHU, E. ZUAZUA.*Allee optimal control of a system in ecology*, in "Mathematical Models and Methods in Applied Sciences", 2018, vol. 28, n$^{o}$ 9, p. 1665–1697, https://hal.archives-ouvertes.fr/hal-01696354

[27] S. XIANG.*Small-time local stabilization for a Korteweg-de Vries equation*, in "Systems and Control Letters", January 2018, vol. Volume 111, p. 64-69, https://hal.archives-ouvertes.fr/hal-01723178

[28] C. ZHANG.*Internal controllability of systems of semilinear coupled one-dimensional wave equations with one control*, in "SIAM Journal on Control and Optimization", July 2018, vol. 56, n$^{o}$ 4, p. 3092 - 3127 [*DOI :* 10.1137/17M1128885], https://hal.archives-ouvertes.fr/hal-01517461

### International Conferences with Proceedings

[29] J. ORŁOWSKI, A. CHAILLET, M. SIGALOTTI, A. DESTEXHE.*Adaptive scheme for pathological oscillations disruption in a delayed neuronal population model*, in "57th IEEE Conference on Decision and Control", Miami Beach, United States, Proceedings of the 57th IEEE Conference on Decision and Control, December 2018, https://hal-centralesupelec.archives-ouvertes.fr/hal-01956472

### Conferences without Proceedings

[30] J.-M. CORON, F. MARBACH, F. SUEUR, P. ZHANG.*On the controllability of the Navier-Stokes equation in a rectangle, with a little help of a distributed phantom force*, in "Journées EDP", Obernai, France, June 2018, https://hal.archives-ouvertes.fr/hal-01970878

### Scientific Books (or Scientific Book chapters)

[31] E. TRÉLAT.*Stabilization of semilinear PDE's, and uniform decay under discretization*, in "London Mathematical Society Lecture Note series", Evolution equations: long time behavior and control, 2018, vol. 439, https://arxiv.org/abs/1506.05883 , https://hal.archives-ouvertes.fr/hal-01165329

### Scientific Popularization

[32] N. AUGIER, U. BOSCAIN, M. SIGALOTTI.*Control of Quantum Systems by Broken Adiabatic Paths*, Ercim, 2018, https://hal.archives-ouvertes.fr/hal-01879704

### Other Publications

[33] N. AUGIER, U. BOSCAIN, M. SIGALOTTI.*Adiabatic ensemble control of a continuum of quantum systems*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01759830

[34] D. BARILARI, Y. CHITOUR, F. JEAN, D. PRANDI, M. SIGALOTTI.*On the regularity of abnormal minimizers for rank 2 sub-Riemannian structures*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01757343

[35] G. BASTIN, J.-M. CORON.*Exponential stability of PI control for Saint-Venant equations with a friction term*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01956634

[36] G. BASTIN, J.-M. CORON, A. HAYAT, P. SHANG.*Boundary feedback stabilization of hydraulic jumps*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02004457

[37] G. BASTIN, J.-M. CORON, A. HAYAT, P. SHANG.*Exponential boundary feedback stabilization of a shock steady state for the inviscid Burgers equation*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01723361

[38] F. BOAROTTO, M. SIGALOTTI.*Dwell-time control sets and applications to the stability analysis of linear switched systems*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-02012606

[39] R. BONALLI, B. HÉRISSÉ, E. TRÉLAT.*Continuity of Pontryagin extremal with respect to delays in nonlinear optimal control*, November 2018, https://arxiv.org/abs/1805.11990 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01802752

[40] Y. CHITOUR, G. MAZANTI, M. SIGALOTTI.*On the gap between deterministic and probabilistic joint spectral radii for discrete-time linear systems*, December 2018, https://arxiv.org/abs/1812.08399 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01961003

[41] J.-M. CORON, A. HAYAT.*PI controllers for 1-D nonlinear transport equation*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01766261

[42] J.-M. CORON, F. MARBACH, F. SUEUR, P. ZHANG.*Controllability of the Navier-Stokes equation in a rectangle with a little help of a distributed phantom force*, 2018, https://arxiv.org/abs/1801.01860 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01676663

[43] J.-M. CORON, H.-M. NGUYEN.*Optimal time for the controllability of linear hyperbolic systems in one dimensional space*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01952134

[44] J.-M. CORON, S. XIANG.*Small-time global stabilization of the viscous Burgers equation with three scalar controls*, March 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01723188

[45] V. FRANCESCHI, G. STEFANI.*Symmetric double bubbles in the Grushin plane*, January 2018, https://arxiv.org/abs/1801.00314 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01674525

[46] A. HAYAT.*Exponential stability of general 1-D quasilinear systems with source terms for the C 1 norm under boundary conditions*, February 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01613139

[47] A. HAYAT.*On boundary stability of inhomogeneous 2 × 2 1-D hyperbolic systems for the $C^1$ norm*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01790104

[48] A. HAYAT.*PI controller for the general Saint-Venant equations*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01827988

[49] A. HAYAT, P. SHANG.*A quadratic Lyapunov function for Saint-Venant equations with arbitrary friction and space-varying slope*, February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01704710

[50] E. HUMBERT, Y. PRIVAT, E. TRÉLAT.*Geometric and spectral characterization of Zoll manifolds, invariant measures and quantum limits*, 2018, https://arxiv.org/abs/1811.12717 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01939709

[51] M. KOHLI.*A metric interpretation of the geodesic curvature in the Heisenberg group*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01916425

[52] A. OLIVIER, C. POUCHOL.*Combination of direct methods and homotopy in numerical optimal control: application to the optimization of chemotherapy in cancer*, January 2018, https://arxiv.org/abs/1707.08038 - working paper or preprint, https://hal-auf.archives-ouvertes.fr/hal-01568779

[53] B. PICCOLI, N. POURADIER DUTEIL, E. TRÉLAT.*Sparse control of Hegselmann-Krause models: Black hole and declustering*, February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01699261

[54] C. POUCHOL, E. TRÉLAT, E. ZUAZUA.*Phase portrait control for 1D monostable and bistable reaction-diffusion equations*, May 2018, https://arxiv.org/abs/1805.10786 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01800382

[55] Y. PRIVAT, E. TRÉLAT, E. ZUAZUA.*Spectral shape optimization for the Neumann traces of the Dirichlet-Laplacian eigenfunctions*, September 2018, https://arxiv.org/abs/1809.05316 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01872896

[56] E. TRÉLAT, G. WANG, Y. XU.*Characterization by observability inequalities of controllability and stabilization properties*, November 2018, https://arxiv.org/abs/1811.01543 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01911941

[57] S. XIANG.*Null controllability of a linearized Korteweg-de Vries equation by backstepping approach*, September 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01468750

[58] C. ZHANG.*Internal rapid stabilization of a 1-D linear transport equation with a scalar feedback*, October 2018, https://arxiv.org/abs/1810.11214 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01905098

[59] C. ZHANG.*Finite-time internal stabilization of a linear 1-D transport equation*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01980349

## References in notes

[60] D. BARILARI, U. BOSCAIN, M. SIGALOTTI (editors). *Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. 1*, EMS Series of Lectures in Mathematics, European Mathematical Society (EMS), Zürich, 2016, vi+324, Lecture notes from the IHP Trimester held at the Institut Henri Poincaré, Paris and from the CIRM Summer School “Sub-Riemannian Manifolds: From Geodesics to Hypoelliptic Diffusion” held in Luminy, Fall 2014, https://doi.org/10.4171/163

[61] D. BARILARI, U. BOSCAIN, M. SIGALOTTI (editors). *Geometry, analysis and dynamics on sub-Riemannian manifolds. Vol. II*, EMS Series of Lectures in Mathematics, European Mathematical Society (EMS), Zürich, 2016, viii+299, Lecture notes from the IHP Trimester held at the Institut Henri Poincaré, Paris and from the CIRM Summer School “Sub-Riemannian Manifolds: From Geodesics to Hypoelliptic Diffusion” held in Luminy, Fall 2014, https://doi.org/10.4171/163

[62] R. ADAMI, U. BOSCAIN.*Controllability of the Schrödinger Equation via Intersection of Eigenvalues*, in "Proceedings of the 44th IEEE Conference on Decision and Control", 2005, p. 1080–1085

[63] A. A. AGRACHEV.*Some open problems*, in "Geometric control theory and sub-Riemannian geometry", Springer INdAM Ser., Springer, Cham, 2014, vol. 5, p. 1–13

[64] A. AGRACHEV, D. BARILARI, L. RIZZI.*Curvature: a variational approach*, in "Mem. Amer. Math. Soc.", 2018, vol. 256, n$^{o}$ 1225, v+142

[65] A. A. AGRACHEV, Y. BARYSHNIKOV, D. LIBERZON.*On robust Lie-algebraic stability conditions for switched linear systems*, in "Systems Control Lett.", 2012, vol. 61, n$^{o}$ 2, p. 347–353, https://doi.org/10.1016/j.sysconle.2011.11.016

[66] A. AGRACHEV, U. BOSCAIN, J.-P. GAUTHIER, F. ROSSI.*The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups*, in "J. Funct. Anal.", 2009, vol. 256, n$^{o}$ 8, p. 2621–2655, https://doi.org/10.1016/j.jfa.2009.01.006

[67] A. AGRACHEV, P. W. Y. LEE.*Generalized Ricci curvature bounds for three dimensional contact subriemannian manifolds*, in "Math. Ann.", 2014, vol. 360, n$^{o}$ 1-2, p. 209–253, https://doi.org/10.1007/s00208-014-1034-6

[68] A. A. AGRACHEV, Y. L. SACHKOV.*Control theory from the geometric viewpoint*, Encyclopaedia of Mathematical Sciences, Springer-Verlag, Berlin, 2004, vol. 87, xiv+412, Control Theory and Optimization, II, https://doi.org/10.1007/978-3-662-06404-7

[69] L. AMBROSIO, N. GIGLI, G. SAVARÉ.*Metric measure spaces with Riemannian Ricci curvature bounded from below*, in "Duke Math. J.", 2014, vol. 163, n$^{o}$ 7, p. 1405–1490, https://doi.org/10.1215/00127094-2681605

[70] L. AMBROSIO, P. TILLI.*Topics on analysis in metric spaces*, Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, 2004, vol. 25, viii+133

[71] M. BALDE, U. BOSCAIN, P. MASON.*A note on stability conditions for planar switched systems*, in "Internat. J. Control", 2009, vol. 82, n$^{o}$ 10, p. 1882–1888, https://doi.org/10.1080/00207170902802992

[72] F. BAUDOIN, N. GAROFALO.*Curvature-dimension inequalities and Ricci lower bounds for sub-Riemannian manifolds with transverse symmetries*, in "J. Eur. Math. Soc. (JEMS)", 2017, vol. 19, n$^{o}$ 1, p. 151–219, https://doi.org/10.4171/JEMS/663

[73] F. BAUDOIN, J. WANG.*Curvature dimension inequalities and subelliptic heat kernel gradient bounds on contact manifolds*, in "Potential Anal.", 2014, vol. 40, n$^{o}$ 2, p. 163–193, https://doi.org/10.1007/s11118-013-9345-x

[74] T. BAYEN.*Analytical parameterization of rotors and proof of a Goldberg conjecture by optimal control theory*, in "SIAM J. Control Optim.", 2008, vol. 47, n$^{o}$ 6, p. 3007–3036 [*DOI :* 10.1137/070705325]

[75] K. BEAUCHARD, J.-M. CORON.*Controllability of a quantum particle in a moving potential well*, in "J. Funct. Anal.", 2006, vol. 232, n$^{o}$ 2, p. 328–389

[76] M. BENAÏM, S. LE BORGNE, F. MALRIEU, P.-A. ZITT.*Qualitative properties of certain piecewise deterministic Markov processes*, in "Ann. Inst. Henri Poincaré Probab. Stat.", 2015, vol. 51, n$^{o}$ 3, p. 1040–1075, https://doi.org/10.1214/14-AIHP619

[77] B. BERRET, C. DARLOT, F. JEAN, T. POZZO, C. PAPAXANTHIS, J. P. GAUTHIER.*The inactivation principle: mathematical solutions minimizing the absolute work and biological implications for the planning of arm movements*, in "PLoS Comput. Biol.", 2008, vol. 4, n$^{o}$ 10, e1000194, 25, https://doi.org/10.1371/journal.pcbi.1000194

[78] F. BLANCHINI.*Nonquadratic Lyapunov functions for robust control*, in "Automatica J. IFAC", 1995, vol. 31, n$^{o}$ 3, p. 451–461, http://dx.doi.org/10.1016/0005-1098(94)00133-4

[79] F. BLANCHINI, S. MIANI.*A new class of universal Lyapunov functions for the control of uncertain linear systems*, in "IEEE Trans. Automat. Control", 1999, vol. 44, n$^{o}$ 3, p. 641–647, http://dx.doi.org/10.1109/9.751368

[80] A. BONFIGLIOLI, E. LANCONELLI, F. UGUZZONI.*Stratified Lie groups and potential theory for their sub-Laplacians*, Springer Monographs in Mathematics, Springer, Berlin, 2007, xxvi+800

[81] M. BORN, V. FOCK.*Beweis des adiabatensatzes*, in "Zeitschrift für Physik A Hadrons and Nuclei", 1928, vol. 51, n$^{o}$ 3–4, p. 165–180

[82] U. BOSCAIN.*Stability of planar switched systems: the linear single input case*, in "SIAM J. Control Optim.", 2002, vol. 41, n$^{o}$ 1, p. 89–112 [*DOI :* 10.1137/S0363012900382837]

[83] U. BOSCAIN, M. CAPONIGRO, T. CHAMBRION, M. SIGALOTTI.*A weak spectral condition for the controllability of the bilinear Schrödinger equation with application to the control of a rotating planar molecule*, in "Comm. Math. Phys.", 2012, vol. 311, n$^{o}$ 2, p. 423–455, https://doi.org/10.1007/s00220-012-1441-z

[84] U. BOSCAIN, M. CAPONIGRO, M. SIGALOTTI.*Multi-input Schrödinger equation: controllability, tracking, and application to the quantum angular momentum*, in "J. Differential Equations", 2014, vol. 256, n$^{o}$ 11, p. 3524–3551, https://doi.org/10.1016/j.jde.2014.02.004

[85] U. BOSCAIN, G. CHARLOT, M. SIGALOTTI.*Stability of planar nonlinear switched systems*, in "Discrete Contin. Dyn. Syst.", 2006, vol. 15, n$^{o}$ 2, p. 415–432, https://doi.org/10.3934/dcds.2006.15.415

[86] U. BOSCAIN, R. A. CHERTOVSKIH, J. P. GAUTHIER, A. O. REMIZOV.*Hypoelliptic diffusion and human vision: a semidiscrete new twist*, in "SIAM J. Imaging Sci.", 2014, vol. 7, n$^{o}$ 2, p. 669–695 [*DOI :* 10.1137/130924731]

[87] U. BOSCAIN, F. CHITTARO, P. MASON, M. SIGALOTTI.*Adiabatic control of the Schroedinger equation via conical intersections of the eigenvalues*, in "IEEE Trans. Automat. Control", 2012, vol. 57, n$^{o}$ 8, p. 1970–1983

[88] U. BOSCAIN, J. DUPLAIX, J.-P. GAUTHIER, F. ROSSI.*Anthropomorphic image reconstruction via hypoelliptic diffusion*, in "SIAM J. Control Optim.", 2012, vol. 50, n$^{o}$ 3, p. 1309–1336 [*DOI :* 10.1137/11082405X]

[89] M. S. BRANICKY.*Multiple Lyapunov functions and other analysis tools for switched and hybrid systems*, in "IEEE Trans. Automat. Control", 1998, vol. 43, n$^{o}$ 4, p. 475–482, Hybrid control systems, https://doi.org/10.1109/9.664150

[90] R. W. BROCKETT.*System theory on group manifolds and coset spaces*, in "SIAM J. Control", 1972, vol. 10, p. 265–284

[91] F. BULLO, A. D. LEWIS.*Geometric control of mechanical systems*, Texts in Applied Mathematics, Springer-Verlag, New York, 2005, vol. 49, xxiv+726, Modeling, analysis, and design for simple mechanical control systems [*DOI :* 10.1007/978-1-4899-7276-7]

[92] C. CARATHÉODORY.*Untersuchungen über die Grundlagen der Thermodynamik*, in "Math. Ann.", 1909, vol. 67, n$^{o}$ 3, p. 355–386, https://doi.org/10.1007/BF01450409

[93] E. CARTAN.*Sur la represéntation géométrique des systèmes matériels non holonomes*, in "Proceedings of the International Congress of Mathematicians. Volume 4.", 1928, p. 253–261

[94] T. CHAMBRION, P. MASON, M. SIGALOTTI, U. BOSCAIN.*Controllability of the discrete-spectrum Schrödinger equation driven by an external field*, in "Ann. Inst. H. Poincaré Anal. Non Linéaire", 2009, vol. 26, n$^{o}$ 1, p. 329–349, https://doi.org/10.1016/j.anihpc.2008.05.001

[95] G. CITTI, A. SARTI.*A cortical based model of perceptual completion in the roto-translation space*, in "J. Math. Imaging Vision", 2006, vol. 24, n$^{o}$ 3, p. 307–326, http://dx.doi.org/10.1007/s10851-005-3630-2

[96] F. COLONIUS, G. MAZANTI.*Decay rates for stabilization of linear continuous-time systems with random switching*, in "Math. Control Relat. Fields", 2019

[97] J.-M. CORON.*Control and nonlinearity*, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007, vol. 136, xiv+426

[98] J.-M. CORON.*On the controllability of nonlinear partial differential equations*, in "Proceedings of the International Congress of Mathematicians. Volume I", Hindustan Book Agency, New Delhi, 2010, p. 238–264

[99] J.-M. CORON.*Global asymptotic stabilization for controllable systems without drift*, in "Math. Control Signals Systems", 1992, vol. 5, n$^{o}$ 3, p. 295–312, https://doi.org/10.1007/BF01211563

[100] D. D'ALESSANDRO.*Introduction to quantum control and dynamics*, Chapman & Hall/CRC Applied Mathematics and Nonlinear Science Series, Chapman & Hall/CRC, Boca Raton, FL, 2008, xiv+343

[101] W. P. DAYAWANSA, C. F. MARTIN.*A converse Lyapunov theorem for a class of dynamical systems which undergo switching*, in "IEEE Trans. Automat. Control", 1999, vol. 44, n$^{o}$ 4, p. 751–760, http://dx.doi.org/10.1109/9.754812

[102] R. DUITS, E. FRANKEN.*Left-invariant diffusions on the space of positions and orientations and their application to crossing-preserving smoothing of HARDI images*, in "Int. J. Comput. Vis.", 2011, vol. 92, n$^{o}$ 3, p. 231–264, https://doi.org/10.1007/s11263-010-0332-z

[103] M. FLIESS, J. LÉVINE, P. MARTIN, P. ROUCHON.*Flatness and defect of non-linear systems: introductory theory and examples*, in "Internat. J. Control", 1995, vol. 61, n$^{o}$ 6, p. 1327–1361, https://doi.org/10.1080/00207179508921959

[104] J. FOISY, M. ALFARO, J. BROCK, N. HODGES, J. ZIMBA.*The standard double soap bubble in* $\mathbf{R}^2$ *uniquely minimizes perimeter*, in "Pacific J. Math.", 1993, vol. 159, n$^{o}$ 1, p. 47–59, http://projecteuclid.org/euclid.pjm/1102634378

[105] A. FRANCI, R. SEPULCHRE.*A three-scale model of spatio-temporal bursting*, in "SIAM J. Appl. Dyn. Syst.", 2016, vol. 15, n$^{o}$ 4, p. 2143–2175 [*DOI :* 10.1137/15M1046101]

[106] S. J. GLASER, U. BOSCAIN, T. CALARCO, C. P. KOCH, W. KÖCKENBERGER, R. KOSLOFF, I. KUPROV, B. LUY, S. SCHIRMER, T. SCHULTE-HERBRÜGGEN, D. SUGNY, F. K. WILHELM.*Training Schrödinger's cat: quantum optimal control. Strategic report on current status, visions and goals for research in Europe*, in "European Physical Journal D", 2015, vol. 69, 279 [*DOI :* 10.1140/EPJD/E2015-60464-1]

[107] E. HAKAVUORI, E. LE DONNE.*Non-minimality of corners in subriemannian geometry*, in "Invent. Math.", 2016, p. 1–12 [*DOI :* 10.1007/s00222-016-0661-9]

[108] R. K. HLADKY, S. D. PAULS.*Minimal surfaces in the roto-translation group with applications to a neuro-biological image completion model*, in "J. Math. Imaging Vision", 2010, vol. 36, n$^{o}$ 1, p. 1–27, https://doi.org/10.1007/s10851-009-0167-9

[109] D. HUBEL, T. WIESEL.*Brain and Visual Perception: The Story of a 25-Year Collaboration*, Oxford University Press, Oxford, 2004

[110] V. JURDJEVIC.*Geometric control theory*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 1997, vol. 52, xviii+492

[111] V. JURDJEVIC, H. J. SUSSMANN.*Control systems on Lie groups*, in "J. Differential Equations", 1972, vol. 12, p. 313–329, https://doi.org/10.1016/0022-0396(72)90035-6

[112] M. KEYL, R. ZEIER, T. SCHULTE-HERBRUEGGEN.*Controlling Several Atoms in a Cavity*, in "New J. Phys.", 2014, vol. 16, 065010

[113] F. KÜSTERS, S. TRENN.*Switch observability for switched linear systems*, in "Automatica J. IFAC", 2018, vol. 87, p. 121–127, https://doi.org/10.1016/j.automatica.2017.09.024

[114] E. LE DONNE, G. P. LEONARDI, R. MONTI, D. VITTONE.*Extremal Curves in Nilpotent Lie Groups*, in "Geom. Funct. Anal.", jul 2013, vol. 23, n$^{o}$ 4, p. 1371–1401 [*DOI :* 10.1007/s00039-013-0226-7], http://arxiv.org/abs/1207.3985

[115] Z. LEGHTAS, A. SARLETTE, P. ROUCHON.*Adiabatic passage and ensemble control of quantum systems*, in "Journal of Physics B", 2011, vol. 44, n$^{o}$ 15

[116] C. LI, I. ZELENKO.*Jacobi equations and comparison theorems for corank 1 sub-Riemannian structures with symmetries*, in "J. Geom. Phys.", 2011, vol. 61, n$^{o}$ 4, p. 781–807, https://doi.org/10.1016/j.geomphys.2010.12.009

[117] D. LIBERZON, J. P. HESPANHA, A. S. MORSE.*Stability of switched systems: a Lie-algebraic condition*, in "Systems Control Lett.", 1999, vol. 37, n$^{o}$ 3, p. 117–122, https://doi.org/10.1016/S0167-6911(99)00012-2

[118] D. LIBERZON.*Switching in systems and control*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 2003, xiv+233, https://doi.org/10.1007/978-1-4612-0017-8

[119] D. LIBERZON.*Calculus of variations and optimal control theory*, Princeton University Press, Princeton, NJ, 2012, xviii+235, A concise introduction

[120] H. LIN, P. J. ANTSAKLIS.*Stability and stabilizability of switched linear systems: a survey of recent results*, in "IEEE Trans. Automat. Control", 2009, vol. 54, n$^{o}$ 2, p. 308–322, http://dx.doi.org/10.1109/TAC.2008.2012009

[121] W. LIU.*Averaging theorems for highly oscillatory differential equations and iterated Lie brackets*, in "SIAM J. Control Optim.", 1997, vol. 35, n$^{o}$ 6, p. 1989–2020 [*DOI :* 10.1137/S0363012994268667]

[122] J. LOTT, C. VILLANI.*Ricci curvature for metric-measure spaces via optimal transport*, in "Ann. of Math. (2)", 2009, vol. 169, n$^{o}$ 3, p. 903–991, https://doi.org/10.4007/annals.2009.169.903

[123] P. MASON, U. BOSCAIN, Y. CHITOUR.*Common polynomial Lyapunov functions for linear switched systems*, in "SIAM J. Control Optim.", 2006, vol. 45, n$^{o}$ 1, p. 226–245 (electronic) [*DOI :* 10.1137/040613147]

[124] P. MASON, M. SIGALOTTI.*Generic controllability properties for the bilinear Schrödinger equation*, in "Comm. Partial Differential Equations", 2010, vol. 35, n$^{o}$ 4, p. 685–706, https://doi.org/10.1080/03605300903540919

[125] L. MASSOULIÉ.*Stability of distributed congestion control with heterogeneous feedback delays*, in "IEEE Trans. Automat. Control", 2002, vol. 47, n$^{o}$ 6, p. 895–902, Special issue on systems and control methods for communication networks, https://doi.org/10.1109/TAC.2002.1008356

[126] M. I. MILLER, A. TROUVÉ, L. YOUNES.*Geodesic shooting for computational anatomy*, in "J. Math. Imaging Vision", 2006, vol. 24, n$^{o}$ 2, p. 209–228, https://doi.org/10.1007/s10851-005-3624-0

[127] M. MIRRAHIMI.*Lyapunov control of a quantum particle in a decaying potential*, in "Ann. Inst. H. Poincaré Anal. Non Linéaire", 2009, vol. 26, n$^{o}$ 5, p. 1743–1765, https://doi.org/10.1016/j.anihpc.2008.09.006

[128] A. P. MOLCHANOV, Y. S. PYATNITSKIY.*Lyapunov functions that specify necessary and sufficient conditions for absolute stability of nonlinear nonstationary control systems, I, II, III*, in "Automat. Remote Control", 1986, vol. 47, p. 344–354, 443–451, 620–630

[129] R. MONTGOMERY.*A tour of subriemannian geometries, their geodesics and applications*, Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2002, vol. 91, xx+259

[130] R. MONTI.*The regularity problem for sub-Riemannian geodesics*, in "Geometric control theory and sub-Riemannian geometry", Springer INdAM Ser., Springer, Cham, 2014, vol. 5, p. 313–332, https://doi.org/10.1007/978-3-319-02132-4_18

[131] R. MONTI, D. MORBIDELLI.*Isoperimetric inequality in the Grushin plane*, in "J. Geom. Anal.", 2004, vol. 14, n$^{o}$ 2, p. 355–368, https://doi.org/10.1007/BF02922077

[132] R. M. MURRAY, S. S. SASTRY.*Nonholonomic motion planning: steering using sinusoids*, in "IEEE Trans. Automat. Control", 1993, vol. 38, n$^{o}$ 5, p. 700–716, https://doi.org/10.1109/9.277235

[133] G. NENCIU.*On the adiabatic theorem of quantum mechanics*, in "J. Phys. A", 1980, vol. 13, n$^{o}$ 2, p. L15–L18, http://stacks.iop.org/0305-4470/13/L15

[134] V. NERSESYAN.*Growth of Sobolev norms and controllability of the Schrödinger equation*, in "Comm. Math. Phys.", 2009, vol. 290, n$^{o}$ 1, p. 371–387

[135] D. PATINO, M. BÂJA, P. RIEDINGER, H. CORMERAIS, J. BUISSON, C. IUNG.*Alternative control methods for DC-DC converters: an application to a four-level three-cell DC-DC converter*, in "Internat. J. Robust Nonlinear Control", 2011, vol. 21, n$^{o}$ 10, p. 1112–1133, https://doi.org/10.1002/rnc.1651

[136] J. PETITOT.*Neurogéomètrie de la vision. Modèles mathématiques et physiques des architectures fonctionnelles*, Les Éditions de l'École Polythechnique, 2008

[137] J. RUESS, J. LYGEROS.*Moment-based methods for parameter inference and experiment design for stochastic biochemical reaction networks*, in "ACM Trans. Model. Comput. Simul.", 2015, vol. 25, n$^{o}$ 2, Art. 8, 25, https://doi.org/10.1145/2688906

[138] A. SARTI, G. CITTI, J. PETITOT.*The symplectic structure of the primary visual cortex*, in "Biol. Cybernet.", 2008, vol. 98, n$^{o}$ 1, p. 33–48, http://dx.doi.org/10.1007/s00422-007-0194-9

[139] H. SCHÄTTLER, U. LEDZEWICZ.*Geometric optimal control*, Interdisciplinary Applied Mathematics, Springer, New York, 2012, vol. 38, xx+640, Theory, methods and examples, https://doi.org/10.1007/978-1-4614-3834-2

[140] H. SCHÄTTLER, U. LEDZEWICZ.*Optimal control for mathematical models of cancer therapies*, Interdisciplinary Applied Mathematics, Springer, New York, 2015, vol. 42, xix+496, An application of geometric methods, https://doi.org/10.1007/978-1-4939-2972-6

[141] R. SHORTEN, F. WIRTH, O. MASON, K. WULFF, C. KING.*Stability criteria for switched and hybrid systems*, in "SIAM Rev.", 2007, vol. 49, n$^{o}$ 4, p. 545–592 [*DOI :* 10.1137/05063516X]

[142] S. SOLMAZ, R. SHORTEN, K. WULFF, F. Ó CAIRBRE.*A design methodology for switched discrete time linear systems with applications to automotive roll dynamics control*, in "Automatica J. IFAC", 2008, vol. 44, n$^{o}$ 9, p. 2358–2363, https://doi.org/10.1016/j.automatica.2008.01.014

[143] E. D. SONTAG.*Input to state stability: basic concepts and results*, in "Nonlinear and optimal control theory", Lecture Notes in Math., Springer, Berlin, 2008, vol. 1932, p. 163–220, https://doi.org/10.1007/978-3-540-77653-6_3

[144] K.-T. STURM.*On the geometry of metric measure spaces. I*, in "Acta Math.", 2006, vol. 196, n$^{o}$ 1, p. 65–131, https://doi.org/10.1007/s11511-006-0002-8

[145] K.-T. STURM.*On the geometry of metric measure spaces. II*, in "Acta Math.", 2006, vol. 196, n$^{o}$ 1, p. 133–177, https://doi.org/10.1007/s11511-006-0003-7

[146] Z. SUN, S. S. GE, T. H. LEE.*Controllability and reachability criteria for switched linear systems*, in "Automatica J. IFAC", 2002, vol. 38, n$^{o}$ 5, p. 775–786, https://doi.org/10.1016/S0005-1098(01)00267-9

[147] Z. SUN, S. S. GE.*Stability theory of switched dynamical systems*, Communications and Control Engineering Series, Springer, London, 2011, xx+253, https://doi.org/10.1007/978-0-85729-256-8

[148] H. J. SUSSMANN.*A regularity theorem for minimizers of real-analytic subriemannian metrics*, in "53rd IEEE Conference on Decision and Control", 2014, p. 4801-4806

[149] K. TAN, X. YANG.*Subriemannian geodesics of Carnot groups of step 3*, in "ESAIM Control Optim. Calc. Var.", 2013, vol. 19, n$^{o}$ 1, p. 274–287, https://doi.org/10.1051/cocv/2012006

[150] S. TEUFEL.*Adiabatic perturbation theory in quantum dynamics*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2003, vol. 1821, vi+236

[151] A. TROUVÉ, L. YOUNES.*Metamorphoses through Lie group action*, in "Found. Comput. Math.", 2005, vol. 5, n$^{o}$ 2, p. 173–198, https://doi.org/10.1007/s10208-004-0128-z

[152] E. TRÉLAT.*Contrôle optimal*, Mathématiques Concrètes. [Concrete Mathematics], Vuibert, Paris, 2005, vi+246, Théorie & applications. [Theory and applications]

[153] M. TUCSNAK, G. WEISS.*Observation and control for operator semigroups*, Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks], Birkhäuser Verlag, Basel, 2009, xii+483, https://doi.org/10.1007/978-3-7643-8994-9

[154] G. TURINICI.*On the controllability of bilinear quantum systems*, in "Mathematical models and methods for ab initio Quantum Chemistry", M. DEFRANCESCHI, C. LE BRIS (editors), Lecture Notes in Chemistry, Springer, 2000, vol. 74

[155] M. VIANA.*Lectures on Lyapunov exponents*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2014, vol. 145, xiv+202, https://doi.org/10.1017/CBO9781139976602

[156] R. VINTER.*Optimal control*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 2000, xviii+507

[157] D. WISNIACKI, G. MURGIDA, P. TAMBORENEA.*Quantum control using diabatic and adiabatic transitions*, in "AIP Conference Proceedings", AIP, 2007, vol. 963, n$^{o}$ 2, p. 840–842

[158] L. YATSENKO, S. GUÉRIN, H. JAUSLIN.*Topology of adiabatic passage*, in "Phys. Rev. A", 2002, vol. 65, 043407, 7

[159] A. VAN DER SCHAFT, H. SCHUMACHER.*An introduction to hybrid dynamical systems*, Lecture Notes in Control and Information Sciences, Springer-Verlag London, Ltd., London, 2000, vol. 251, xiv+174, https://doi.org/10.1007/BFb0109998

# Project-Team CASCADE

## Construction and Analysis of Systems for Confidentiality and Authenticity of Data and Entities

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Algorithmics, Computer Algebra and Cryptology**

# Table of contents

# Project-Team CASCADE

*Creation of the Project-Team: 2008 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A4. - Security and privacy
A4.3. - Cryptography
A4.3.1. - Public key cryptography
A4.3.3. - Cryptographic protocols
A4.8. - Privacy-enhancing technologies
A7. - Theory of computation
A8.5. - Number theory

**Other Research Topics and Application Domains:**

B6.4. - Internet of things
B9.5.1. - Computer science
B9.10. - Privacy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

David Pointcheval [Team leader, CNRS, Researcher, HDR]
Michel Ferreira Abdalla [CNRS, Researcher, HDR]
Georg Fuchsbauer [Inria, Researcher]
Brice Minaud [Inria, Researcher, from Oct 2018]
Hoeteck Wee [CNRS, Researcher, HDR]

**PhD Students**

Balthazar Bauer [Inria]
Jérémy Chotard [CNRS]
Aurélien Dupin [Thales]
Pierre-Alain Dupont [DGA, until Aug 2018]
Romain Gay [Ecole Normale Supérieure Paris]
Dahmun Goudarzi [CryptoExperts, until Sep 2018]
Chloé Hébant [CNRS]
Louiza Khati [ANSSI]
Michele Minelli [Ecole Normale Supérieure Paris, until Oct 2018]
Anca Nitulescu [CNRS]
Michele Orrù [CNRS]
Antoine Plouviez [Inria, from Sep 2018]
Razvan Rosie [Ecole Normale Supérieure Paris]
Mélissa Rossi [Thales]
Quentin Santos [Orange Labs]
Quoc Huy Vu [Ecole Normale Supérieure Paris, from Oct 2018]

**Post-Doctoral Fellow**

Pooya Farshim [Ecole Normale Supérieure Paris]

**Administrative Assistant**

Nathalie Gaudechoux [Inria]

# 2. Overall Objectives

## 2.1. Presentation

Cryptographic algorithms are the equivalent of locks, seals, security stamps and identification documents over the Internet. They are essential to protect our online bank transactions, credit cards, medical and personal information, and to support e-commerce and e-government. They come in different flavors. Encryption algorithms are necessary to protect sensitive information such as medical data, financial information and Personal Identification Numbers (PINs) from prying eyes. Digital signature algorithms (in combination with hash functions) and MAC algorithms replace hand-written signatures in electronic transactions. Identification protocols allow to securely verify the identity of a remote party. As a whole, cryptology is a research area with a high strategic impact in industry, for individuals, and for society as a whole. The research activity of project-team CASCADE addresses the following topics, which cover most of the areas that are currently active in the international cryptographic community, with a focus on public-key algorithms:

1. Implementation of cryptographic algorithms, and applied cryptography;
2. Algorithm and protocol design, and provable security;
3. Theoretical and practical attacks.

## 2.2. Design of Provably Secure Primitives and Protocols

Since the beginning of public-key cryptography, with the seminal Diffie-Hellman paper, many suitable algorithmic problems for cryptography have been proposed and many cryptographic schemes have been designed, together with more or less heuristic proofs of their security relative to the intractability of the underlying problems. However, many of those schemes have thereafter been broken. The simple fact that a cryptographic algorithm withstood cryptanalytic attacks for several years has often been considered as a kind of validation procedure, but schemes may take a long time before being broken. An example is the Chor-Rivest cryptosystem, based on the knapsack problem, which took more than 10 years to be totally broken by Serge Vaudenay, whereas before this attack it was believed to be strongly secure. As a consequence, the lack of attacks at some time should never be considered as a full security validation of the proposal.

A completely different paradigm is provided by the concept of “provable” security. A significant line of research has tried to provide proofs in the framework of computational complexity theory (a.k.a. “reductionist” security proofs): the proofs provide reductions from a well-studied problem (factoring, RSA or the discrete logarithm) to an attack against a cryptographic protocol.

At the beginning, researchers just tried to define the security notions required by actual cryptographic schemes, and then to design protocols which could achieve these notions. The techniques were directly derived from complexity theory, providing polynomial reductions. However, their aim was essentially theoretical. They were indeed trying to minimize the required assumptions on the primitives (one-way functions or permutations, possibly trapdoor, etc), without considering practicality. Therefore, they just needed to design a scheme with polynomial-time algorithms, and to exhibit polynomial reductions from the basic mathematical assumption on the hardness of the underlying problem to an attack of the security notion, in an asymptotic way. However, such a result has no practical impact on actual security. Indeed, even with a polynomial reduction, one may be able to break the cryptographic protocol within a few hours, whereas the reduction just leads to an algorithm against the underlying problem which requires many years. Therefore, those reductions only prove the security when very huge (and thus maybe unpractical) parameters are in use, under the assumption that no polynomial-time algorithm exists to solve the underlying problem. For many years, more efficient reductions have been expected, under the denomination of either “exact security" or “concrete security”, which provide more practical security results, with concrete efficiency properties.

Unfortunately, in many cases, even just provable security is at the cost of an important loss in terms of efficiency for the cryptographic protocol. Thus, some models have been proposed, trying to deal with the security of efficient schemes: some concrete objects are identified with ideal (or black-box) ones. For example, it is by now usual to identify hash functions with ideal random functions, in the so-called "random-oracle model". Similarly, block ciphers are identified with families of truly random permutations in the "ideal cipher model". Another kind of idealization has also been introduced in cryptography, the black-box group, where the group operation, in any algebraic group, is defined by a black-box: a new element necessarily comes from the addition (or the subtraction) of two already known elements. It is by now called the "generic group model", extended to the bilinear and multi-linear setting. Some works even require several ideal models together to provide some new validations.

But still, such idealization cannot be instantiated in practice, and so one prefers provable security without such idealized assumptions, under new and possibly stronger computational assumptions. As a consequence, a cryptographer has to deal with the following four important steps, which are **all** main goals of ours:

**computational assumptions**, which are the foundation of the security. We thus need to have a strong evidence that the computational problems are reasonably hard to solve.

**security model**, which makes precise the security notions one wants to achieve, as well as the means the adversary may be given. We contribute to this point, in several ways:

- by providing security models for many primitives and protocols;
- by enhancing some classical security models;
- by considering new means for the adversary, such as side-channel information.

**design** of new schemes/protocols, or more efficient ones, with additional features, etc.

**security proof**, which consists in exhibiting a reduction.

# 3. Research Program

## 3.1. Quantum-Safe Cryptography

The security of almost all public-key cryptographic protocols in use today relies on the presumed hardness of problems from number theory such as factoring and computing discrete logarithms. This is problematic because these problems have very similar underlying structure, and its unforeseen exploit can render all currently used public-key cryptography insecure. This structure was in fact exploited by Shor to construct efficient quantum algorithms that break all hardness assumptions from number theory that are currently in use. And so naturally, an important area of research is to build provably secure protocols based on mathematical problems that are unrelated to factoring and discrete log. One of the most promising directions in this line of research is using lattice problems as a source of computational hardness, which also offer features that other alternative public-key cryptosystems (such as MQ-based, code-based or hash-based schemes) cannot provide.

## 3.2. Advanced Encryption

Fully Homomorphic Encryption (FHE) has become a very active research area since 2009, when IBM announced the discovery of a FHE scheme by Craig Gentry. FHE allows to perform any computation on encrypted data, yielding the result encrypted under the same key. This enables outsourcing computation in the Cloud, on encrypted data, so the Cloud provider does not learn any information. However, FHE does not allow to share the result.

Functional encryption is another recent tool that allows an authority to deliver functional decryption keys, for any function $f$ of his choice, so that when applied to the encryption of a message $m$, the functional decryption key yields $f(m)$. Since $m$ can be a large vector, $f$ can be an aggregation or statistical function: on encrypted data, one can get the result $f(m)$ in clear.

While this functionality has initially been defined in theory, our team has been very active in designing concrete instantiations for practical purposes.

## 3.3. Security amidst Concurrency on the Internet

Cryptographic protocols that are secure when executed in isolation can become completely insecure when multiple such instances are executed concurrently (as is unavoidable on the Internet) or when used as a part of a larger protocol. For instance, a man-in-the-middle attacker participating in two simultaneous executions of a cryptographic protocol might use messages from one of the executions in order to compromise the security of the second – Lowe's attack on the Needham-Schroeder authentication protocol and Bleichenbacher's attack on SSL work this way. Our research addresses security amidst concurrent executions in secure computation and key exchange protocols.

Secure computation allows several mutually distrustful parties to collaboratively compute a public function of their inputs, while providing the same security guarantees as if a trusted party had performed the computation. Potential applications for secure computation include anonymous voting, privacy-preserving auctions and data-mining. Our recent contributions on this topic include

1. new protocols for secure computation in a model where each party interacts only once, with a single centralized server; this model captures communication patterns that arise in many practical settings, such as that of Internet users on a website, and
2. efficient constructions of universally composable commitments and oblivious transfer protocols, which are the main building blocks for general secure computation.

In key exchange protocols, we are actively involved in designing new password-authenticated key exchange protocols, as well as the analysis of the widely-used SSL/TLS protocols.

## 3.4. Electronic Currencies and the Blockchain

Electronic cash (e-cash) was first proposed in the 1980s but has never been deployed on a large scale. Other means of digital payments are instead largely replacing physical cash, but they do not respect the citizens' right to privacy, which includes their right of anonymous payments of moderate sums. Recently, so-called decentralized currencies, such as Bitcoin, have become a third type of payments in addition to physical cash, and card and other (non-anonymous) electronic payments. The continuous growth of popularity and usage of this new kind of currencies, also called "cryptocurrencies", have triggered a renewed interest in cryptographic e-cash.

On the one hand, our group investigates "centralized" e-cash, in keeping with the current economic model that has money be issued by (central) banks (while cryptocurrencies use money distribution as an incentive for participation in the system, on which its stability hinges). Of particular interest among centralized e-cash schemes is transferable e-cash, which allows users to transfer coins between each other without interacting with a third party (or the blockchain). Existing efficient e-cash schemes are not transferable, as they require coins to be deposited at the bank after having been used in a payment. Our goal is to propose efficient transferable e-cash schemes.

Another direction concerns (decentralized) cryptocurrencies, whose adoption has grown tremendously over the last few years. While in Bitcoin all transactions are publicly posted on the so-called "blockchain", other cryptocurrencies such as *Zcash* respect user privacy, whose security guarantees we have analyzed. Apart from privacy, two pressing challenges for cryptocurrencies, and blockchains in general, are sustainability and scalability. Regarding the former, we are addressing the electricity waste caused by the concept of "proof of work" used by all major cryptocurrencies by proposing alternatives; for the latter, we are working on proposals that avoid the need for all data having to be stored on the blockchain forever.

Blockchains have meanwhile found many other applications apart from electronic money. Together with Microsoft Research, our group investigates decentralized means of authentication that uses cryptography to guarantee privacy.

# 4. Application Domains

## 4.1. Privacy for the Cloud

Many companies have already started the migration to the Cloud and many individuals share their personal informations on social networks. While some of the data are public information, many of them are personal and even quite sensitive. Unfortunately, the current access mode is purely right-based: the provider first authenticates the client, and grants him access, or not, according to his rights in the access-control list. Therefore, the provider itself not only has total access to the data, but also knows which data are accessed, by whom, and how: privacy, which includes secrecy of data (confidentiality), identities (anonymity), and requests (obliviousness), should be enforced. Moreover, while high availability can easily be controlled, and thus any defect can immediately be detected, failures in privacy protection can remain hidden for a long time. The industry of the Cloud introduces a new implicit trust requirement: nobody has any idea at all of where and how his data are stored and manipulated, but everybody should blindly trust the providers. The providers will definitely do their best, but this is not enough. Privacy-compliant procedures cannot be left to the responsibility of the provider: however strong the trustfulness of the provider may be, any system or human vulnerability can be exploited against privacy. This presents too huge a threat to tolerate. *The distribution of the data and the secrecy of the actions must be given back to the users. It requires promoting privacy as a global security notion.*

In order to protect the data, one needs to encrypt it. Unfortunately, traditional encryption systems are inadequate for most applications involving big, complex data. Recall that in traditional public key encryption, a party encrypts data to a single known user, which lacks the expressiveness needed for more advanced data sharing. In enterprise settings, a party will want to share data with groups of users based on their credentials. Similarly, individuals want to selectively grant access to their personal data on social networks as well as documents and spreadsheets on Google Docs. Moreover, the access policy may even refer to users who do not exist in the system at the time the data is encrypted. Solving this problem requires an entirely new way of encrypting data.

A first natural approach would be **fully homomorphic encryption** (FHE, see above), but a second one is also **functional encryption**, that is an emerging paradigm for public-key encryption: it enables more fine-grained access control to encrypted data, for instance, the ability to specify a decryption policy in the ciphertext so that only individuals who satisfy the policy can decrypt, or the ability to associate keywords to a secret key so that it can only decrypt documents containing the keyword. Our work on functional encryption centers around two goals:

1. to obtain more efficient pairings-based functional encryption;
2. and to realize new functionalities and more expressive functional encryption schemes.

Another approach is **secure multi-party computation protocols**, where interactivity might provide privacy in a more efficient way. Recent implicit interactive proofs of knowledge can be a starting point. But stronger properties are first expected for improving privacy. They can also be integrated into new ad-hoc broadcast systems, in order to distribute the management among several parties, and eventually remove any trust requirements.

Strong privacy for the Cloud would have a huge societal impact since it would revolutionize the trust model: users would be able to make safe use of outsourced storage, namely for personal, financial and medical data, without having to worry about failures or attacks of the server.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

- Melissa Rossi received a 2018 Google's WomenTechmakers Scholarship.

# 6. New Results

## 6.1. Results

All the results of the team have been published in journals or conferences (see the list of publications). They are all related to the research program (see before) and the research projects (see after):

- Advanced primitives for privacy in the cloud
- Efficient functional encryption
- Several predicate-encryption schemes
- New primitives for efficient anonymous authentication
- Analyses of currently deployed zero-knowledge SNARKs

# 7. Partnerships and Cooperations

## 7.1. National Initiatives with Industry

### *7.1.1. CryptoComp*

Program: FUI

Duration: October 2014 – November 2018

Coordinator: CryptoExperts

Partners: CEA, CNRS, Kalray, Inria, Dictao, Université de Limoges, VIACESS, Bertin technologies, GEMALTO

Local coordinator: David Pointcheval

We aim at studying delegation of computations to the cloud, in a secure way.

### *7.1.2. ANBLIC*

Title: Analysis in Blind Clouds

Program: FUI

Duration: January 2018 – December 2020

Coordinator: Wallix

Partners: UPEC, CEA, Ingenico, Atos, SOGETI, CoeSSI

Local coordinator: David Pointcheval

The main goal is to industrialize for the first time several privacy enhancing technologies that are on the edge of theory and practice.

Fully Homomorphic Encryption let cloud providers compute arbitrary functions on their client's encrypted data, ensuring at the same time full privacy and functionality. Functional Encryption is a refinement of classical encryption, which allows data owners to delegate fine-grained access to their data. Thus it is possible to enable the computation of aggregated statistics over your personal data, while cryptographically ensuring its confidentiality.

However both these technologies still suffer from prohibitive inefficiencies for business applications. ANBLIC's academic partners will create new cryptographic schemes and performance models, tailored for industrial use cases, and create the first real-life scenario of encrypted queries on encrypted data and on open data.

### *7.1.3. RISQ*

Program: GDN

Duration: February 2017 – September 2020

Coordinator: Secure-IC

Partners: ANSSI, AIRBUS, C-S, CEA LIST, CryptoExperts, Inria/ENS/CASCADE, GEMALTO, Inria POLSYS, Inria AriC, IRISA, Orange Labs, THALES, UVSQ, PCQC

Local coordinator: Michel Abdalla

The main goal of RISQ is to help the French Industry and Academia become a significant international player in the transition to post-quantum cryptography.

## 7.2. National Collaborations with Academics

### *7.2.1. EnBiD*

Title: Encryption for Big Data

Program: ANR JCJC

Duration: October 2014 – September 2019

PI: Hoeteck Wee

Partners: Université Paris 2, Université Limoges

The main objective of this project is to study techniques for efficient and expressive functional encryption schemes. Functional encryption is a novel paradigm for public-key encryption that enables both fine-grained access control and selective computation on encrypted data, as is necessary to protect big, complex data in the cloud.

### *7.2.2. EfTrEC*

Title: Efficient Transferable E-Cash

Program: ANR JCJC

Duration: October 2016 – September 2020

PI: Georg Fuchsbauer

Partners: Université Paris 2

This project deals with e-cash systems which let users transfer electronic coins between them offline. The main objectives of this project are:

- establish a clean formal model for the primitive;
- construct schemes which are practically efficient;
- develop schemes that are resistant to attacks on quantum computers.

### *7.2.3. ALAMBIC*

Title: AppLicAtions of MalleaBIlity in Cryptography

Program: ANR PRC

Duration: October 2016 – September 2020

PI: Damien Vergnaud

Partners: ENS Lyon, Université Limoges

The main objectives of the proposal are the following:

- Define theoretical models for “malleable” cryptographic primitives that capture strong practical attacks (in particular, in the settings of secure computation outsourcing, server-aided cryptography, cloud computing and cryptographic proof systems);

- Analyze the security and efficiency of primitives and constructions that rely on malleability;
- Conceive novel cryptographic primitives and constructions (for secure computation outsourcing, server-aided cryptography, multi-party computation, homomorphic encryption and their applications);
- Implement these new constructions in order to validate their efficiency and effective security.

## 7.3. European Initiatives

### *7.3.1. CryptoAction*

Title: Cryptography for Secure Digital Interaction

Program: H2020 ICT COST

Duration: April 2014 – April 2018

Local coordinator: Michel Abdalla

The aim of this COST CryptoAction is to stimulate interaction between the different national efforts in order to develop new cryptographic solutions and to evaluate the security of deployed algorithms with applications to the secure digital interactions between citizens, companies and governments.

### *7.3.2. CryptoCloud*

Title: Cryptography for the Cloud

Program: FP7 ERC Advanced Grant

Duration: June 2014 – May 2020

PI: David Pointcheval

The goal of the CryptoCloud project is to develop new interactive tools to provide privacy in the Cloud.

### *7.3.3. SAFEcrypto*

Title: Secure Architectures of Future Emerging Cryptography

Program: H2020

Duration: January 2015 – January 2019

Coordinator: The Queen's University of Belfast

Partners: Inria/ENS (France), Emc Information Systems International (Ireland), Hw Communications (United Kingdom), The Queen's University of Belfast (United Kingdom), Ruhr-Universitaet Bochum (Germany), Thales Uk (United Kingdom), Universita della Svizzera italiana (Switzerland), IBM Research Zurich (Switzerland)

Local coordinator: Michel Abdalla

SAFEcrypto will provide a new generation of practical, robust and physically secure post quantum cryptographic solutions that ensure long-term security for future ICT systems, services and applications. Novel public-key cryptographic schemes (digital signatures, authentication, public-key encryption, identity-based encryption) will be developed using lattice problems as the source of computational hardness. The project will involve algorithmic and design optimisations, and implementations of the lattice-based cryptographic schemes addressing the cost, energy consumption, performance and physical robustness needs of resource-constrained applications, such as mobile, battery-operated devices, and of real-time applications such as network security, satellite communications and cloud. Currently a significant threat to cryptographic applications is that the devices on which they are implemented leak information, which can be used to mount attacks to recover secret

information. In SAFEcrypto the first analysis and development of physical-attack resistant methodologies for lattice-based cryptographic implementations will be undertaken. Effective models for the management, storage and distribution of the keys utilised in the proposed schemes (key sizes may be in the order of kilobytes or megabytes) will also be provided. This project will deliver proof-of-concept demonstrators of the novel lattice-based public-key cryptographic schemes for three practical real-word case studies with real-time performance and low power consumption requirements. In comparison to current state-of-the-art implementations of conventional public-key cryptosystems (RSA and Elliptic Curve Cryptography (ECC)), SAFEcrypto's objective is to achieve a range of lattice-based architectures that provide comparable area costs, a 10-fold speed-up in throughput for real-time application scenarios, and a 5-fold reduction in energy consumption for low-power and embedded and mobile applications.

### *7.3.4. ECRYPT-NET*

Title: Advanced Cryptographic Technologies for the Internet of Things and the Cloud

Program: H2020 ITN

Duration: March 2015 – February 2019

Coordinator: KU Leuven (Belgium)

Partners: KU Leuven (Belgium), Inria/ENS (France), Ruhr-Universität Bochum (Germany), Royal Holloway, University of London (UK), University of Bristol (UK), CryptoExperts (France), NXP Semiconductors (Belgium), Technische Universiteit Eindhoven (the Netherlands)

Local coordinator: Michel Abdalla

ECRYPT-NET is a research network of six universities and two companies, as well as 7 associated companies, that intends to develop advanced cryptographic techniques for the Internet of Things and the Cloud and to create efficient and secure implementations of those techniques on a broad range of platforms.

### *7.3.5. aSCEND*

Title: Secure Computation on Encrypted Data

Program: H2020 ERC Starting Grant

Duration: June 2015 – May 2020

PI: Hoeteck Wee

The goals of the aSCEND project are (i) to design pairing- and lattice-based functional encryption that are more efficient and ultimately viable in practice; and (ii) to obtain a richer understanding of expressive functional encryption schemes and to push the boundaries from encrypting data to encrypting software.

### *7.3.6. FENTEC*

Title: Functional Encryption Technologies

Program: H2020

Duration: January 2018 – December 2020

Coordinator: ATOS Spain SA

Scientific coordinator: Michel Abdalla

Partners: Inria/ENS (France), Flensburg University (Germany), KU Leuven (Belgium), University of Helsinki (Finland), Nagra (Switzerland), XLAB (Switzerland), University of Edinburgh (United Kingdom), WALLIX (France)

Local coordinator: Michel Abdalla

Functional encryption (FE) has recently been introduced as a new paradigm of encryption systems to overcome all-or-nothing limitations of classical encryption. In an FE system the decryptor deciphers a function over the message plaintext: such functional decryptability makes it feasible to process encrypted data (e.g. on the Internet) and obtain a partial view of the message plaintext. This extra flexibility over classical encryption is a powerful enabler for many emerging security technologies (i.e. controlled access, searching and computing on encrypted data, program obfuscation...). FENTEC's mission is to make the functional encryption paradigm ready for wide-range applications, integrating it in ICT technologies as naturally as classical encryption. The primary objective is the efficient and application-oriented development of functional encryption systems. FENTEC's team of cryptographers, software and hardware experts and information technology industry partners will document functional encryption needs of specific applications and subsequently design, develop, implement and demonstrate applied use of functional cryptography. Ultimately, a functional encryption library for both SW and HW-oriented application will be documented and made public so that it may be used by European ICT entities. With it, the FENTEC team will build emerging security technologies that increase the trustworthiness of the European ICT services and products. Concretely, the FENTEC team will showcase the expressiveness and versatility of the functional encryption paradigm in 3 use cases:

- Privacy-preserving digital currency, enforcing flexible auditing models
- Anonymous data analytics enabling computation of statistics over encrypted data, protecting European Fundamental Rights of Data Protection and Privacy
- Key and content distribution with improved performance & efficiency as foundational technology for establishing secure communication among a vast number of IOT devices.

## 7.4. International Initiatives with Industry

### *7.4.1. CryptBloC*

Title: Cryptography for the Blockchain

Partners: MSR Redmond (USA), MSR Cambridge (UK), Inria

Duration: October 2017 – October 2021

PI: Georg Fuchsbauer

The goal of this Microsoft-Inria joint project on privacy and decentralization is to use cryptography to improve privacy on the blockchain and decentralized systems more generally. We will investigate means of privacy-preserving authentication, such as electronic currencies, and other applications of blockchain and distributed transparency mechanisms.

## 7.5. International Research Visitors

- Yuval Ishai (Technion)
- Dan Boneh (Stanford)
- Katsuyuki Takashima (Mitsubishi and Kyushu University)
- Tal Malkin (Columbia)
- Adam O'Neill (Georgetown University)
- Julian Loss (Ruhr Universität Bochum)

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. Events and Activities*

- punctual seminars are organized: https://crypto.di.ens.fr/web2py/index/seminars
- quarterly Paris Crypto Days (https://pariscryptoday.github.io) supported by CryptoCloud and aSCEND
- BibTeX database of papers related to Cryptography, open and widely used by the community (https://cryptobib.di.ens.fr)
- LATCA Bertinoro workshop (http://crypto-events.di.ens.fr/LATCA/)

*8.1.1.2. Steering Committees of International Conferences*

- steering committee of CANS: David Pointcheval
- steering committee of PKC: David Pointcheval
- steering committee of LATINCRYPT: Michel Abdalla (chair)
- steering committee of PAIRING: Michel Abdalla

*8.1.1.3. Board of International Organisations*

- Board of the *International Association for Cryptologic Research* (IACR): Michel Abdalla (2013 – 2018)

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Program Committee Member*

- CT-RSA '18 – 16-20 April (San Francisco, California, USA): David Pointcheval
- Eurocrypt '18 – 29 April-3 May (Tel Aviv, Israel): Georg Fuchsbauer and David Pointcheval
- Crypto '18 – 19-23 August (Santa Barbara, USA): Georg Fuchsbauer and Hoeteck Wee
- SCN '18 – 5-7 September (Amalfi, Italy): Georg Fuchsbauer and Romain Gay
- TCC '18 – 11-14 November (Goa, India): Hoeteck Wee

### *8.1.3. Editorial Boards of Journals*

Editor-in-Chief

- of the *International Journal of Applied Cryptography (IJACT)* – Inderscience Publishers: David Pointcheval

Associate Editor

- of *IET Information Security*: Michel Abdalla
- of *ETRI Journal*: Michel Abdalla
- of *Applicable Algebra in Engineering, Communication and Computing*: David Pointcheval

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

- Master: Michel Abdalla, David Pointcheval, Cryptography, M2, MPRI
- Master: David Pointcheval, Cryptography, M2, ESIEA
- Bachelor: Georg Fuchsbauer, David Pointcheval, Jacques Stern, Hoeteck Wee, Introduction to Cryptology, L3/M1, ENS
- Bachelor: Georg Fuchsbauer, Cryptology, 3rd year, ESGI

### *8.2.2. Defenses*

- PhD: Raphaël Bost, Algorithmes de recherche sur bases de données chiffrées, Univ. Rennes I, January 8th, 2018 (Supervisors: Pierre-Alain Fouque & David Pointcheval)
- PhD: Rafael Del Pino, Efficient Lattice-Based ZeroKnowledge Proofs And Applications, ENS, June 1st, 2018 (Supervisors: Vadim Lyubashevsky & David Pointcheval)

- PhD: Pierre-Alain Dupont, Advanced password-authenticated key exchanges, ENS, August 29th, 2018 (Supervisor: David Pointcheval)
- PhD: Dahmun Goudarzi, Implémentations Sécurisées de Chiffrement par Bloc contre les Attaques Physiques, ENS, September 21st, 2018 (Supervisor: Damien Vergnaud)
- PhD: Michele Minelli, Fully Homomorphic Encryption for Machine Learning, ENS, October 26th, 2018 (Supervisors: Michel Abdalla & Hoeteck Wee)
- PhD: Quentin Santos, Cryptography for Pragmatic Distributed Trust and the Role of Blockchain, ENS, December 20th, 2018 (Supervisor: David Pointcheval)

### *8.2.3. Supervision*

- PhD in progress: Aurélien Dupin, Multi-Party Computations, from 2015, David Pointcheval (with Christophe Bidan, at Rennes)
- PhD in progress: Romain Gay, Functional Encryption, from 2015, Michel Abdalla & Hoeteck Wee
- PhD in progress: Louiza Khati, Disk Encryption Modes, from 2015, Damien Vergnaud
- PhD in progress: Anca Nitulescu, Verifiable Outsourced Computations, from 2015, Michel Abdalla & David Pointcheval
- PhD in progress: Razvan Rosie, Practical Functional Encryption Schemes For the Cloud, from 2015, Michel Abdalla & Hoeteck Wee
- PhD in progress: Jérémy Chotard, Attribute-Based Encryption, from 2016, David Pointcheval (with Duong Hieu Phan, at Limoges)
- PhD in progress: Michele Orrù, Functional Encryption, from 2016, Hoeteck Wee & Georg Fuchsbauer
- PhD in progress: Balthazar Bauer, Transferable e-Cash, from 2017, Georg Fuchsbauer
- PhD in progress: Chloé Hébant, Big Data and Privacy, from 2017, David Pointcheval (with Duong Hieu Phan, at Limoges)
- PhD in progress: Mélissa Rossi, Post-Quantum Cryptography, from 2017, Michel Abdalla (with Henri Gilbert at ANSSI and Thomas Prest at Thales)
- PhD in progress: Antoine Plouviez, Privacy and Decentralization, from 2018, Georg Fuchsbauer
- PhD in progress: Quoc Huy Vu, Quantum Cryptography, from 2018, Céline Chevalier

### *8.2.4. Committees*

- PhD Raphaël Bost. *Algorithmes de recherche sur bases de données chiffrées* – Université Rennes I – France – January 8th, 2018: David Pointcheval (Co-supervisor)
- PhD Xavier Bultel. *Mécanismes de délégation pour les primitives de cryptographie à clé publique* – Université Clermont Auvergne – France – May 17th, 2018: David Pointcheval (President)
- PhD Luca Nizzardo. *Cryptographic Techniques for the Security of Cloud and Blockchain Systems* – IMDEA / UPM – Spain – May 24th, 2018: Georg Fuchsbauer (Reviewer)
- PhD Rafael Del Pino. *La cryptographie à base de réseaux* – Ecole Normale Supérieure – France – June 1st, 2018: David Pointcheval (Co-supervisor)
- PhD Pierre-Alain Dupont. *Advanced password-authenticated key exchanges* – Ecole Normale Supérieure – France – August 29th, 2018: David Pointcheval (Supervisor)
- PhD Dahmun Goudarzi. *Implémentations Sécurisées de Chiffrement par Bloc contre les Attaques Physiques* – Ecole Normale Supérieure – France – September 21st, 2018: David Pointcheval (President)
- PhD Cédric Van Rompay. *Protocoles Multi-Utilisateurs de Recherche sur Bases de Données Chiffrées* – Eurecom, Sophia Antipolis, Télécom ParisTech – France – October 4th, 2018: David Pointcheval (President)

- PhD Damien Ligier. *Functional encryption applied to privacy-preserving classification: practical use, performances and security* – CEA, Saclay, IMT Atlantique – France – October 15th, 2018: David Pointcheval (President)
- PhD Fabrice Mouhartem. *Cryptographie protégeant la vie privée à base de couplages et de réseaux* – ENS Lyon – France – October 18th, 2018: David Pointcheval (Reviewer)
- PhD Michele Minelli. *Fully Homomorphic Encryption for Machine Learning* – Ecole Normale Supérieure – France – October 26th, 2018: Michel Abdalla and Hoeteck Wee (Co-supervisors)
- PhD Nadim Kobeissi. *Formal Verification for Real-World Cryptographic Protocols and Implementations* – Inria Paris – France – December 10th, 2018: David Pointcheval (President)
- PhD Quentin Santos. *Cryptography for Pragmatic Distributed Trust and the Role of Blockchain* – Ecole Normale Supérieure – France – December 20th, 2018: David Pointcheval (Supervisor)

# 9. Bibliography

## Major publications by the team in recent years

[1] M. ABDALLA, D. CATALANO, D. FIORE.*Verifiable Random Functions: Relations to Identity-Based Key Encapsulation and New Constructions*, in "Journal of Cryptology", 2014, vol. 27, n$^{o}$ 3, p. 544-593

[2] M. ABE, G. FUCHSBAUER, J. GROTH, K. HARALAMBIEV, M. OHKUBO.*Structure-Preserving Signatures and Commitments to Group Elements*, in "Journal of Cryptology", 2016, vol. 29, n$^{o}$ 2, p. 363–421

[3] F. BENHAMOUDA, O. BLAZY, C. CHEVALIER, D. POINTCHEVAL, D. VERGNAUD.*New Techniques for SPHFs and Efficient One-Round PAKE Protocols*, in "Advances in Cryptology – Proceedings of CRYPTO '13 (1)", R. CANETTI, J. A. GARAY (editors), Lecture Notes in Computer Science, Springer, 2013, vol. 8042, p. 449-475

[4] P. CHAIDOS, V. CORTIER, G. FUCHSBAUER, D. GALINDO.*BeleniosRF: A Non-interactive Receipt-Free Electronic Voting Scheme*, in "Proceedings of the 23rd ACM Conference on Computer and Communications Security (CCS '16)", E. R. WEIPPL, S. KATZENBEISSER, C. KRUEGEL, A. C. MYERS, S. HALEVI (editors), ACM Press, 2016, p. 1614–1625

[5] I. DINUR, O. DUNKELMAN, N. KELLER, A. SHAMIR.*New Attacks on Feistel Structures with Improved Memory Complexities*, in "Advances in Cryptology – Proceedings of CRYPTO '15 (1)", R. GENNARO, M. ROBSHAW (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9215, p. 433-454

[6] Y. DODIS, D. POINTCHEVAL, S. RUHAULT, D. VERGNAUD, D. WICHS.*Security Analysis of Pseudo-Random Number Generators with Input: /dev/random is not Robust*, in "Proceedings of the 20th ACM Conference on Computer and Communications Security (CCS '13)", Berlin, Germany, V. D. GLIGOR, M. YUNG (editors), ACM Press, 2013, p. 647–658

[7] R. GAY, D. HOFHEINZ, E. KILTZ, H. WEE.*Tightly CCA-Secure Encryption Without Pairings*, in "Advances in Cryptology – Proceedings of Eurocrypt '16 (2)", M. FISCHLIN, J.-S. CORON (editors), Lecture Notes in Computer Science, Springer, 2016, vol. 9665, p. 1–27

[8] S. GORBUNOV, V. VAIKUNTANATHAN, H. WEE.*Predicate Encryption for Circuits from LWE*, in "Advances in Cryptology – Proceedings of CRYPTO '15 (2)", R. GENNARO, M. ROBSHAW (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9216, p. 503-523

[9] V. LYUBASHEVSKY, C. PEIKERT, O. REGEV.*On Ideal Lattices and Learning with Errors over Rings*, in "Journal of the ACM", 2013, vol. 60, n° 6, p. 43:1–43:35

[10] W. QUACH, H. WEE, D. WICHS.*Laconic Function Evaluation and Applications*, in "59th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2018)", M. THORUP (editor), IEEE, 2018

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[11] P.-A. DUPONT.*Advanced password-authenticated key exchanges*, PSL Research University, August 2018, https://hal.inria.fr/tel-01868828

[12] D. GOUDARZI.*Secure Implementation of Block Ciphers against Physical Attacks*, ENS Paris - Ecole Normale Supérieure de Paris, September 2018, https://hal.inria.fr/tel-01896103

[13] M. MINELLI.*Fully Homomorphic Encryption for Machine Learning*, PSL University, October 2018, https://hal.archives-ouvertes.fr/tel-01918263

[14] Q. SANTOS.*Cryptography for Pragmatic Distributed Trust and the Role of Blockchain*, PSL Research University ; École Normale Supérieure, December 2018, https://hal.archives-ouvertes.fr/tel-01966109

## Articles in International Peer-Reviewed Journal

[15] M. ABDALLA, M. BELLARE, G. NEVEN.*Robust Encryption*, in "Journal of Cryptology", April 2018, vol. 31, n° 2, p. 307-350 [*DOI :* 10.1007/s00145-017-9258-8], https://hal.archives-ouvertes.fr/hal-01538113

[16] M. ABDALLA, F. BENHAMOUDA, A. PASSELÈGUE, K. PATERSON.*Related-Key Security for Pseudorandom Functions Beyond the Linear Barrier*, in "Journal of Cryptology", October 2018, vol. 31, n° 4, p. 917-964 [*DOI :* 10.1007/s00145-017-9274-8], https://hal.inria.fr/hal-01723012

[17] M. ABDALLA, F. BENHAMOUDA, D. POINTCHEVAL.*On the Tightness of Forward-Secure Signature Reductions*, in "Journal of Cryptology", February 2018, p. 1-67 [*DOI :* 10.1007/s00145-018-9283-2], https://hal.inria.fr/hal-01722996

[18] S. CANARD, D. H. PHAN, D. POINTCHEVAL, V. C. TRINH.*A new technique for compacting ciphertext in multi-channel broadcast encryption and attribute-based encryption*, in "Theoretical Computer Science", May 2018, vol. 723, p. 51 - 72 [*DOI :* 10.1016/J.TCS.2018.02.036], https://hal.inria.fr/hal-01903749

[19] G. FUCHSBAUER, C. HANSER, D. SLAMANIG.*Structure-Preserving Signatures on Equivalence Classes and Constant-Size Anonymous Credentials*, in "Journal of Cryptology", 2018 [*DOI :* 10.1007/s00145-018-9281-4], https://hal.inria.fr/hal-01870052

## International Conferences with Proceedings

[20] M. ABDALLA, D. CATALANO, D. FIORE, R. GAY, B. URSU.*Multi-Input Functional Encryption for Inner Products: Function-Hiding Realizations and Constructions without Pairings*, in "Advances in Cryptology – CRYPTO 2018", Santa Barbara, United States, H. SHACHAM, A. BOLDYREVA (editors), Lecture Notes in Computer Science, August 2018, vol. 10991 [*DOI :* 10.1007/978-3-319-96884-1_20], https://hal.archives-ouvertes.fr/hal-01900273

[21] M. BARBOSA, P. FARSHIM.*Indifferentiable Authenticated Encryption*, in "Advances in Cryptology – CRYPTO 2018", Santa Barbara, United States, H. SHACHAM, A. BOLDYREVA (editors), LNCS - Lecture Notes in Computer Science, Springer, August 2018, vol. 10991 [*DOI :* 10.1007/978-3-319-96884-1_7], https://hal.inria.fr/hal-01904141

[22] G. BARTHE, S. BELAÏD, T. ESPITAU, P.-A. FOUQUE, B. GRÉGOIRE, M. ROSSI, M. TIBOUCHI.*Masking the GLP Lattice-Based Signature Scheme at Any Order*, in "Eurocrypt 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques", Tel Aviv, Israel, J. B. NIELSE, V. RIJME (editors), Lecture Notes in Computer Science, Springer, April 2018, vol. 10821, p. 354-384 [*DOI :* 10.1007/978-3-319-78375-8_12], https://hal.inria.fr/hal-01900708

[23] B. BAUER, P. FARSHIM, S. MAZAHERI.*Combiners for Backdoored Random Oracles*, in "Advances in Cryptology – CRYPTO 2018", Santa Barbara, United States, H. SHACHAM, A. BOLDYREVA (editors), LNCS - Lecture Notes in Computer Science, Springer, August 2018, vol. 10992 [*DOI :* 10.1007/978-3-319-96881-0_10], https://hal.inria.fr/hal-01866724

[24] B. BAUER, L. ISENMANN.*Discrete Morse theory for the collapsibility of supremum sections*, in "ICGT: International Colloquium on Graph Theory and combinatorics", Lyon, France, July 2018, https://arxiv.org/abs/1803.09577 , https://hal.archives-ouvertes.fr/hal-01867246

[25] S. CANARD, D. POINTCHEVAL, Q. SANTOS, J. TRAORÉ.*Practical Strategy-Resistant Privacy-Preserving Elections*, in "ESORICS 2018", Barcelona, Spain, Computer Security. ESORICS 2018, Springer, September 2018, vol. Lecture Notes in Computer Science, n$^{o}$ 11099 [*DOI :* 10.1007/978-3-319-98989-1_17], https://hal.inria.fr/hal-01903777

[26] S. CANARD, D. POINTCHEVAL, Q. SANTOS, J. TRAORÉ.*Privacy-Preserving Plaintext-Equality of Low-Entropy Inputs*, in "ACNS 2018 - 16th International Conference on Applied Cryptography and Network Security", Leuven, Belgium, A. PRENEEL, F. VERCAUTERE (editors), Applied Cryptography and Network Security. ACNS 2018, Springer, July 2018, vol. Lecture Notes in Computer Science, n$^{o}$ 10892 [*DOI :* 10.1007/978-3-319-93387-0_14], https://hal.inria.fr/hal-01903746

[27] J. CHEN, J. GONG, L. KOWALCZYK, H. WEE.*Unbounded ABE via Bilinear Entropy Expansion, Revisited*, in "EUROCRYPT 2018 - Annual International Conference on the Theory and Applications of Cryptographic Techniques", Tel Aviv, Israel, J. B. NIELSE, V. RIJME (editors), LNCS - Lecture Notes in Computer Science, Springer, April 2018, vol. 10820, p. 503-534 [*DOI :* 10.1007/978-3-319-78381-9_19], https://hal.inria.fr/hal-01899901

[28] J. CHEN, J. GONG, H. WEE.*Improved Inner-product Encryption with Adaptive Security and Full Attribute-hiding*, in "ASIACRYPT 2018 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, December 2018, https://hal.inria.fr/hal-01900153

[29] Y. CHEN, V. VAIKUNTANATHAN, B. WATERS, H. WEE, D. WICHS.*Traitor-Tracing from LWE Made Simple and Attribute-Based*, in "TCC 2018 - Theory of Cryptography Conference", Goa, India, November 2018, https://hal.inria.fr/hal-01900152

[30] Y. CHEN, V. VAIKUNTANATHAN, H. WEE.*GGH15 Beyond Permutation Branching Programs: Proofs, Attacks, and Candidates*, in "CRYPTO 2018 - 38th Annual International Cryptology Conference", Santa Barbara, United States, H. SHACHAM, A. BOLDYREVA (editors), LNCS - Lecture Notes in Computer Science, Spinger, August 2018, vol. 10991, https://hal.inria.fr/hal-01899903

[31] J. CHOTARD, E. DUFOUR SANS, R. GAY, D. POINTCHEVAL, D. H. PHAN.*Decentralized Multi-Client Functional Encryption for Inner Product*, in "ASIACRYPT '18 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, Advances in Cryptology - ASIACRYPT '18, Springer, December 2018, vol. Lecture Notes in Computer Science, n$^{o}$ 11273 [*DOI :* 10.1007/978-3-030-03329-3_24], https://hal.archives-ouvertes.fr/hal-01668020

[32] G. COUTEAU, A. DUPIN, P. MÉAUX, M. ROSSI, Y. ROTELLA.*On the Concrete Security of Goldreich's Pseudorandom Generator*, in "ASIACRYPT 2018 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, LNCS - Lecture Notes in Computer Science, Springer, December 2018, vol. 11273, p. 96-124 [*DOI :* 10.1007/978-3-030-03329-3_4], https://hal.inria.fr/hal-01944772

[33] A. DUPIN, D. POINTCHEVAL, C. BIDAN.*On the Leakage of Corrupted Garbled Circuits*, in "ProvSec 2018 - 12th International Conference on Provable Security", Jeju, South Korea, J. BAEK, W. SUSILO, J. KIM (editors), Provable Security. ProvSec 2018, Springer, October 2018, vol. Lecture Notes in Computer Science, n$^{o}$ 11192 [*DOI :* 10.1007/978-3-030-01446-9_1], https://hal.inria.fr/hal-01903806

[34] P.-A. DUPONT, J. HESSE, D. POINTCHEVAL, L. REYZIN, S. YAKOUBOV.*Fuzzy Password-Authenticated Key Exchange*, in "EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic", Tel Aviv, Israel, J. B. NIELSE, V. RIJME (editors), Advances in Cryptology – EUROCRYPT 2018, Springer, April 2018, vol. Lecture Notes in Computer Science, n$^{o}$ 10822 [*DOI :* 10.1007/978-3-319-78372-7_13], https://hal.inria.fr/hal-01903733

[35] P. FARSHIM, J. HESSE, D. HOFHEINZ, E. LARRAIA.*Graded Encoding Schemes from Obfuscation*, in "PKC 2018 - 21st IACR International Conference on Practice and Theory of Public-Key Cryptography", Rio De Janeiro, Brazil, M. ABDALLA, R. DAHAB (editors), LNCS - Lecture Notes in Computer Science, Springer, March 2018, vol. 10769 [*DOI :* 10.1007/978-3-319-76581-5_13], https://hal.inria.fr/hal-01904151

[36] G. FUCHSBAUER.*Subversion-Zero-Knowledge SNARKs*, in "PKC 2018 - 21st IACR International Conference on Practice and Theory of Public-Key Cryptography", Rio de Jaineiro, Brazil, M. ABDALLA, R. DAHAB (editors), LNCS - Lecture Notes in Computer Science, Springer, March 2018, vol. 10769, p. 315-347 [*DOI :* 10.1007/978-3-319-76578-5_11], https://hal.inria.fr/hal-01869978

[37] G. FUCHSBAUER, R. GAY.*Weakly Secure Equivalence-Class Signatures from Standard Assumptions*, in "PKC 2018 - 21st IACR International Conference on Practice and Theory of Public-Key Cryptography", Rio de Jaineiro, Brazil, M. ABDALLA, R. DAHAB (editors), Springer, March 2018, p. 153-183 [*DOI :* 10.1007/978-3-319-76581-5_6], https://hal.inria.fr/hal-01869971

[38] G. FUCHSBAUER, E. KILTZ, J. LOSS.*The Algebraic Group Model and its Applications*, in "CRYPTO 2018 - 38th Annual International Cryptology Conference", Santa Barbara, United States, H. SHACHAM, A. BOLDYREVA (editors), LNCS - Lecture Notes in Computer Science, Springer, August 2018, vol. 10992, p. 33-62 [*DOI :* 10.1007/978-3-319-96881-0_2], https://hal.inria.fr/hal-01870015

[39] G. FUCHSBAUER, M. ORRÙ.*Non-interactive Zaps of Knowledge*, in "ACNS 2018 - 16th International Conference on Applied Cryptography and Network Security", Leuven, Belgium, A. PRENEEL, F. VERCAUTERE (editors), Springer, July 2018, vol. LNCS - Lecture notes in computer science, n$^{o}$ 10892, p. 44-62 [*DOI :* 10.1007/978-3-319-93387-0_3], https://hal.inria.fr/hal-01870005

[40] R. GAY, D. HOFHEINZ, L. KOHL, J. PAN.*More Efficient (Almost) Tightly Secure Structure-Preserving Signatures*, in "EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques", Tel Aviv, Israel, J. B. NIELSE, V. RIJME (editors), Springer, April 2018 [*DOI :* 10.1007/978-3-319-78375-8_8], https://hal.archives-ouvertes.fr/hal-01900268

[41] R. GAY, L. KOWALCZYK, H. WEE.*Tight Adaptively Secure Broadcast Encryption with Short Ciphertexts and Keys*, in "SCN 2018: Security and Cryptography for Networks", Amalfi, Italy, September 2018, https://hal.inria.fr/hal-01900275

[42] R. GENNARO, M. MINELLI, A. NITULESCU, M. ORRÙ.*Lattice-Based zk-SNARKs from Square Span Programs*, in "ACM CCS 2018", Toronto, Canada, October 2018, https://hal.archives-ouvertes.fr/hal-01743360

[43] P. GRUBBS, M.-S. LACHARITÉ, B. MINAUD, K. G. PATERSON.*Learning to Reconstruct: Statistical Learning Theory and Encrypted Database Attacks*, in "IEEE Symposium on Security and Privacy (S&P) 2019", San Francisco, United States, May 2019, https://hal.inria.fr/hal-01974962

[44] L. KHATI, D. VERGNAUD.*Analysis and Improvement of an Authentication Scheme in Incremental Cryptography*, in "Selected Areas in Cryptography - SAC 2018", Calgary, Canada, C. CID, M. J. JACOBSON JR. (editors), Lecture Notes in Computer Science, Springer, August 2018, vol. 11349, p. 50-70 [*DOI :* 10.1007/978-3-030-10970-7_3], https://hal.inria.fr/hal-01893905

[45] T. LIU, V. VAIKUNTANATHAN, H. WEE.*Towards Breaking the Exponential Barrier for General Secret Sharing*, in "EUROCRYPT 2018 - 37th Annual International Conference on the Theory and Applications of Cryptographic Techniques", Tel Aviv, Israel, J. B. NIELSE, V. RIJME (editors), LNCS - Lecture Notes in Computer Science, Springer, April 2018, vol. 10820, https://hal.inria.fr/hal-01899902

[46] S. PARK, A. KWON, G. FUCHSBAUER, P. GAŽI, J. ALWEN, K. PIETRZAK.*SpaceMint: A Cryptocurrency Based on Proofs of Space*, in "Financial Cryptography and Data Security 2018", Curaçao, Curaçao, February 2018, https://hal.inria.fr/hal-01869990

[47] D. POINTCHEVAL, O. SANDERS.*Reassessing Security of Randomizable Signatures*, in "CT-RSA 2018 - The Cryptographers' Track at the RSA Conference 2018", San Francisco, United States, Topics in Cryptology – CT-RSA 2018, Springer, April 2018, vol. Lecture Notes in Computer Science, n$^{o}$ 10808 [*DOI :* 10.1007/978-3-319-76953-0_17], https://hal.inria.fr/hal-01903717

[48] W. QUACH, H. WEE, D. WICHS.*Laconic Function Evaluation and Applications*, in "FOCS 2018 - 59th Annual IEEE Symposium on Foundations of Computer Science", Paris, France, October 2018, https://hal.inria.fr/hal-01899904

# Team COML

## The Cognitive Machine Learning Team

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
**Paris**

THEME
**Language, Speech and Audio**

# Table of contents

# Team COML

*Creation of the Team: 2017 May 04*

**Keywords:**

**Computer Science and Digital Science:**

A3.4.2. - Unsupervised learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A5.7. - Audio modeling and processing
A5.7.1. - Sound
A5.7.3. - Speech
A5.7.4. - Analysis
A5.8. - Natural language processing
A6.3.3. - Data processing
A9.2. - Machine learning
A9.3. - Signal analysis
A9.4. - Natural language processing
A9.6. - Decision support
A9.7. - AI algorithmics

**Other Research Topics and Application Domains:**

B1.2. - Neuroscience and cognitive science
B1.2.2. - Cognitive science

# 1. Team, Visitors, External Collaborators

**Research Scientist**

Xuan Nga Cao [Ecole des hautes études en sciences sociales, Researcher]

**Faculty Member**

Emmanuel Dupoux [Ecole des hautes études en sciences sociales, Professor, HDR]

**Technical Staff**

Mathieu Bernard [Inria]
Julien Karadayi [Ecole des hautes études en sciences sociales]
Catherine Urban [Ecole des hautes études en sciences sociales]

**PhD Students**

Maria Julia Carbajal [Ecole Normale Supérieure Paris, until Sep 2018]
Rahma Chaabouni [Ecole Normale Supérieure Paris]
Adriana Carolina Guevara Rukoz [Ecole Normale Supérieure Paris, until Sep 2018]
Elin Larsen [Ecole Normale Supérieure Paris, until Dec 2018]
Rachid Riad [Ecole des hautes études en sciences sociales]
Neil Zeghidour [Facebook]

**Administrative Assistant**

Chantal Chazelas [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

Brain-inspired machine learning algorithms combined with big data have recently reached spectacular results, equalling or beating humans on specific high level tasks (e.g. the game of go). However, there are still a lot of domains in which even humans infants outperform machines: unsupervised learning of rules and language, common sense reasoning, and more generally, cognitive flexibility (the ability to quickly transfer competence from one domain to another one).

The aim of the Cognitive Computing team is to *reverse engineer* such human abilities, i.e., to construct effective and scalable algorithms which perform as well (or better) than humans, when provided with similar data, study their mathematical and algorithmic properties and test their empirical validity as models of humans by comparing their output with behavioral and neuroscientific data. The expected results are more adaptable and autonomous machine learning algorithm for complex tasks, and quantitative models of cognitive processes which can used to predict human developmental and processing data. Most of the work is focused on speech and language and common sense reasoning.

# 3. Research Program

## 3.1. Background

In recent years, Artificial Intelligence (AI) has achieved important landmarks in matching or surpassing human level performance on a number of high level tasks (playing chess and go, driving cars, categorizing picture, etc., [28], [31], [36], [27], [33]). These strong advances were obtained by deploying on large amounts of data, massively parallel learning architectures with simple brain-inspired 'neuronal' elements. However, humans brains still outperform machines in several key areas (language, social interactions, common sense reasoning, motor skills), and are more flexible : Whereas machines require extensive expert knowledge and massive training for each particular application, humans learn autonomously over several time scales: over the developmental scale (months), humans infants acquire cognitive skills with noisy data and little or no expert feedback (weakly/unsupervised learning)[1]; over the short time scale (minutes, seconds), humans combine previously acquired skills to solve new tasks and apply rules systematically to draw inferences on the basis of extremely scarce data (learning to learn, domain adaptation, one- or zero-shot learning) [30].

The general aim of CoML, following the roadmap described in [1], is to bridge the gap in cognitive flexibility between humans and machines learning in language processing and common sense reasoning. We conduct work in three areas: weakly supervised and unsupervised algorithms, datasets and benchmarks, and machine intelligence evaluation.

## 3.2. Weakly/Unsupervised Learning

Much of standard machine learning is construed as regression or classification problems (mapping input data to expert-provided labels). Human infants rarely learn in this fashion, at least before going to school: they learn language, social cognition, and common sense autonomously (without expert labels) and when adults provide feedback, it is ambiguous and noisy and cannot be taken as a gold standard. Modeling or mimicking such achievement requires deploying unsupervised or weakly supervised algorithms which are less well known than their supervised counterparts.

We take inspiration from infant's landmarks during their first years of life: they are able to learn acoustic models, a lexicon, and susbtantive elements of language models and world models from raw sensory inputs. Building on previous work [3], [7], [11], we use DNN and Bayesian architectures to model the emergence of linguistic representations without supervision. Our focus is to establish how the labels in supervised settings can be replaced by weaker signals coming either from multi-modal input or from hierarchically organised linguistic levels.

At the level of phonetic representations, we study how cross-modal information (lips and self feedback from articulation) can supplement top-down lexical information in a weakly supervised setting. We use siamese architectures or Deep CCA algorithms to combine the different views. We study how an attentional framework and uncertainty estimation can flexibly combine these informations in order to adapt to situations where one view is selectively degraded.

At the level of lexical representations, we study how audio/visual parallel information (ie. descriptions of images or activities) can help in segmenting and clustering word forms, and vice versa, help in deriving useful visual features. To achieve this, we will use architectures deployed in image captioning or sequence to sequence translation [34].

At the level of semantic and conceptual representations, we study how it is possible to learn elements of the laws of physics through the observation of videos (object permanence, solidity, spatio-temporal continuity, inertia, etc.), and how objects and relations between objects are mapped onto language.

## 3.3. Evaluating Machine Intelligence

Increasingly, complicated machine learning systems are being incorporated into real-life applications (e.g. self-driving cars, personal assistants), even though they cannot be formally verified, guaranteed statistically, nor even explained. In these cases, a well defined *empirical approach* to evaluation can offer interesting insights into the functioning and offer some control over these algorithms.

Several approaches exist to evaluate the 'cognitive' abilities of machines, from the subjective comparison of human and machine performance [35] to application-specific metrics (e.g., in speech, word error rate). A recent idea consist in evaluating an AI system in terms of it's *abilities* [29] , i.e., functional components within a more global cognitive architecture [32]. Psychophysical testing can offer batteries of tests using simple tasks that are easy to understand by humans or animals (e.g, judging whether two stimuli are same or different, or judging whether one stimulus is 'typical') which can be made selective to a specific component and to rare but difficult or adversarial cases. Evaluations of learning rate, domain adaptation and transfer learning are simple applications of these measures. Psychophysically inspired tests have been proposed for unsupervised speech and language learning [10], [6].

## 3.4. Documenting human learning

Infants learn their first language in a spontaneous fashion, across a lot of variation in amount of speech and the nature of the infant/adult interaction. In some linguistic communities, adults barely address infants until they can themselves speak. Despite these large variations in quantity and content, language learning proceeds at similar paces. Documenting such resilience is an essential step in understanding the nature of the learning algorithms used by human infants. Hence, we propose to collect and/or analyse large datasets of inputs to infants and correlate this with outcome measure (phonetic learning, vocabulary growth, syntactic learning, etc.).

# 4. Application Domains

## 4.1. Speech processing for underresourced languages

We plan to apply our algorithms for the unsupervised discovery of speech units to problems relevant to language documentation and the construction of speech processing pipelines for underresourced languages.

## 4.2. Tools for the analysis of naturalistic speech corpora

Daylong recordings of speech in the wild gives rise a to number of specific analysis difficulties. We plan to use our expertise in speech processing to develop tools for performing signal processing and helping annotation of such resources for the purpose of phonetic or linguistic analysis.

# 5. New Software and Platforms

## 5.1. abkhazia

KEYWORDS: Speech recognition - Speech-text alignment
FUNCTIONAL DESCRIPTION: The Abkhazia sofware makes it easy to obtain simple baselines for supervised ASR (using Kaldi) and ABX tasks (using ABXpy) on the large corpora of speech recordings typically used in speech engineering, linguistics or cognitive science research.

- Contact: Emmanuel Dupoux
- URL: https://github.com/bootphon/abkhazia

## 5.2. TDE

*Term Discovery Evaluation*
KEYWORDS: NLP - Speech recognition - Speech
SCIENTIFIC DESCRIPTION: This toolbox allows the user to judge of the quality of a word discovery algorithm. It evaluates the algorithms on these criteria : - Boundary : efficiency of the algorithm to found the actual boundaries of the words - Group : efficiency of the algorithm to group similar words - Token/Type: efficiency of the algorithm to find all words from the corpus (types), and to find all occurences (token) of these words. - NED : Mean of the edit distance across all the word pairs found by the algorithm - Coverage : efficiency of the algorithm to find every discoverable phone in the corpus
FUNCTIONAL DESCRIPTION: Toolbox to evaluate algorithms that segment speech into words. It allows the user to evaluate the efficiency of algorithms to segment speech into words, and create clusters of similar words.

- Contact: Emmanuel Dupoux
- URL: https://github.com/bootphon/TDE

## 5.3. ABXpy

KEYWORDS: Evaluation - Speech recognition - Machine learning
FUNCTIONAL DESCRIPTION: The ABX package gives a performance score to speech recognition systems by measuring their capacity to discriminate linguistic contrasts (accents, phonemes, speakers, etc...)

- Contact: Emmanuel Dupoux
- URL: https://github.com/bootphon/ABXpy

## 5.4. h5features

KEYWORD: File format
FUNCTIONAL DESCRIPTION: The h5features python package provides easy to use and efficient storage of large features data on the HDF5 binary file format.

- Contact: Emmanuel Dupoux
- URL: https://github.com/bootphon/h5features

# 6. New Results

## 6.1. Speech and Audio Processing from the Raw Waveform

State-of-the-art speech technology systems (e.g., ASR and TTS) rely on fixed, hand-crafted features such as mel-filterbanks to preprocess the waveform before the training pipeline. This is at odds with recent work in machine vision where hand-crafter features (SIFT, etc) have been succesfully replaced by features derived from raw pixels trained jointly with a downstream task. In this line of work, we explored how a similar approach could be undertaken for audio and speech processing.

- In [24], we train a bank of complex filters that operates at the level of the raw speech signal and feeds into a convolutional neural network for phone recognition. These time-domain filterbanks (TD-filterbanks) are initialized as an approximation of MFSC, and then fine-tuned jointly with the remaining convolutional network. We perform phone recognition experiments on TIMIT and show that for several architectures, models trained on TD-filterbanks consistently out-perform their counterparts trained on comparable MFSC. We get our best performance by learning all front-end steps, from pre-emphasis up to averaging. Finally, we observe that the filters at convergence have an asymmetric impulse response while preserving some analyticity.
- In [25], we study end-to-end systems trained directly from the raw waveform, building on two alternatives for trainable replacements of mel-filterbanks that use a convolutional architecture. The first one is inspired by gammatone filterbanks [4], [9], and the second one by the scattering transform [24]. We propose two modifications to these architectures and systematically compare them to mel-filterbanks, on the Wall Street Journal dataset. The first modification is the addition of an instance normalization layer, which greatly improves on the gammatone-based trainable filterbanks and speeds up the training of the scattering-based filterbanks. The second one relates to the low-pass filter used in these approaches. These modifications consistently improve performances for both approaches, and remove the need for a careful initialization in scattering-based trainable filterbanks. In particular, we show a consistent improvement in word error rate of the trainable filterbanks relatively to comparable mel-filterbanks. It is the first time end-to-end models trained from the raw signal significantly outperform mel-filterbanks on a large vocabulary task under clean recording conditions.
- Recent progress in deep learning for audio synthesis opens the way to models that directly produce the waveform, shifting away from the traditional paradigm of relying on vocoders or MIDI synthesizers. Despite their successes, current state-of-the-art neural audio synthesizers such as WaveNet and SampleRNN [12], [8] suffer from prohibitive training and inference times because they are based on autoregressive models that generate audio samples one at a time at a rate of 16kHz. In this work [26], we study the more computationally efficient alternative of generating the waveform frame-by-frame with large strides. We present SING, a lightweight neural audio synthesizer for the original task of generating musical notes given desired instrument, pitch and velocity. Our model is trained end-to-end to generate notes from nearly 1000 instruments with a single decoder, thanks to a new loss function that minimizes the distances between the log spectrograms of the generated and target waveforms. On the generalization task of synthesizing notes for pairs of pitch and instrument not seen during training, SING produces audio with significantly improved perceptual quality compared to a state-of-the-art autoencoder based on WaveNet [4] as measured by a Mean Opinion Score (MOS), and is about 32 times faster for training and 2,500 times faster for inference.

## 6.2. Development of cognitively inspired algorithms

Speech and language processing in humans infants and adults is particularly efficient. We use these as sources of inspiration for developing novel machine learning and speech technology algorithms. In this area, our results are as follows:

- In [22], we summarize the accomplishments of a multi-disciplinary 6-weeks workshop organized by E. Dupoux (PI) at Carnegy Mellon Univerrsity (Pittsburgh), funded through the Jelinek Memorial Summer Workshop Program of Johns Hopkins University. The workshop explored the computational and scientific issues surrounding the discovery of linguistic units (subwords and words) in a language without orthography. We studied the replacement of orthographic transcriptions by images and/or translated text in a well-resourced language to help unsupervised discovery from raw speech.
- Developing speech technologies for low-resource languages has become a very active research field over the last decade. Among others, Bayesian models have shown some promising results on artificial examples but still lack of in situ experiments. In [20] , we apply state-of-the-art Bayesian models to unsupervised Acoustic Unit Discovery (AUD) in a real low-resource language scenario. We also show that Bayesian models can naturally integrate information from other resourceful languages by means of informative prior leading to more consistent discovered units. Finally, discovered acoustic units are used, either as the 1-best sequence or as a lattice, to perform word segmentation. Word segmentation results show that this Bayesian approach clearly outperforms a Segmental-DTW baseline on the same corpus.
- Fixed-length embeddings of words are very useful for a variety of tasks in speech and language processing. In [19], we systematically explore two methods of computing fixed-length embeddings for variable-length sequences. We evaluate their susceptibility to phonetic and speaker-specific variability on English, a high resource language, and Xitsonga, a low resource language, using two evaluation metrics: ABX word discrimination and ROC-AUC on same-different phoneme n-grams. We show that a simple downsampling method supplemented with length information can be competitive with the variable-length input feature representation on both evaluations. Recurrent autoencoders trained without supervision can yield even better results at the expense of increased computational complexity.
- Recent studies have investigated siamese network architectures for learning invariant speech representations using same-different side information at the word level. In [21], we investigate systematically an often ignored component of siamese networks: the sampling procedure (how pairs of same vs. different tokens are selected). We show that sampling strategies taking into account Zipf's Law, the distribution of speakers and the proportions of same and different pairs of words significantly impact the performance of the network. In particular, we show that word frequency compression improves learning across a large range of variations in number of training pairs. This effect does not apply to the same extent to the fully unsupervised setting, where the pairs of same-different words are obtained by spoken term discovery. We apply these results to pairs of words discovered using an unsupervised algorithm and show an improvement on state-of-the-art in unsupervised representation learning using siamese networks.
- Unsupervised spoken term discovery is the task of finding recurrent acoustic patterns in speech without any annotations. Current approaches consists of two steps: (1) discovering similar patterns in speech, and (2) partitioning those pairs of acoustic tokens using graph clustering methods. In, [23] we propose a new approach for the first step. Previous systems used various approximation algorithms to make the search tractable on large amounts of data. Our approach is based on an optimized $k$-nearest neighbours (KNN) search coupled with a fixed word embedding algorithm. The results show that the KNN algorithm is robust across languages, consistently outperforms the DTW-based baseline, and is competitive with current state-of-the-art spoken term discovery systems.

## 6.3. Test of the psychological validity of AI algorithms.

In this section, we focus on the utilisation of machine learning algorithms of speech and language processing to derive testable quantitative predictions in humans (adults or infants).

- Two PhDs were defended this year. In [14], Adriana Guavara Rukoz presented a computational model of the perception of non-native speech contrasts based on standard ASR pipelines is presented. An adaptation of the model is proposed to account for forced-choice classification psycholinguistic

experiments and directly reproduced classical results. The general finding is that, suprisingly, the acoustic model part of a phone recognizer is sufficient to account for experimental data, even those apparently related to phonotactic properties of the native language. The 'language model' part does not improve the correlation with adult data (if anything, it degrades it). Yet the match between model and human is not perfect, and is was hypothetized that improvement in the acoustic model could help. In [13], Julia Maria Carbajal presented a study of the effect of multilingual exposure on language acquisition. She used a computational model of language separation based on i-vectors to reproduce some of the known effects of phonological distance on language discrimination in infants.

- In [16], we investigate whether infant-directed speech (IDS) facilitates lexical learning when compared to adult-directed speech (ADS). To study this, we compare the distinctiveness of the lexicon at two levels, acoustic and phonological, using a large database of spontaneous speech in Japanese. At the acoustic level we show that, as has been documented before for phonemes, the realizations of words are more variable and less discriminable in IDS. At the phonological level, we find that despite a slight increase in the number of phonological neighbors, the IDS lexicon contains more distinctive words (such as onomatopeias). Combining the acoustic and phonological metrics together in a global discrimination score, the two effects cancel each other out and the IDS lexicon winds up being as discriminable as its ADS counterpart. We discuss the implication of these findings for the view of IDS as hyperspeech, i.e., a register whose purpose is to facilitate language acquisition.
- Existing theories of cross-linguistic phonetic category perception agree that listeners perceive foreign sounds by mapping them onto their native phonetic categories. Yet, none of the available theories specify a way to compute this mapping. As a result, they cannot provide systematic quantitative predictions and remain mainly descriptive. Here [17], Automatic Speech Recognition (ASR) systems are used to provide a fully specified mapping between foreign and native sounds. This is shown to provide a quantitative model that can account for several empirically attested effects in human cross-linguistic phonetic category perception.
- Spectacular progress in the information processing sciences (machine learning, wearable sensors) promises to revolutionize the study of cognitive development. In [15], we analyse the conditions under which 'reverse engineering' language development, i.e., building an effective system that mimics infant's achievements, can contribute to our scientific understanding of early language development. We argue that, on the computational side, it is important to move from toy problems to the full complexity of the learning situation, and take as input as faithful reconstructions of the sensory signals available to infants as possible. On the data side, accessible but privacy-preserving repositories of home data have to be setup. On the psycholinguistic side, specific tests have to be constructed to benchmark humans and machines at different linguistic levels. We discuss the feasibility of this approach and present an overview of current results.

## 6.4. Applications and tools for researchers

Some of CoMLs' activity is to produce speech and language technology tools that facilitate research into language development or clinical applications.

- In [18], we present BabyCloud, a platform for capturing, storing and analyzing daylong audio recordings and photographs of children's linguistic environments, for the purpose of studying infant's cognitive and linguistic development and interactions with the environment. The proposed platform connects two communities of users: families and academics, with strong innovation potential for each type of users. For families, the platform offers a novel functionality: the ability for parents to follow the development of their child on a daily basis through language and cognitive metrics (growth curves in number of words, verbal complexity, social skills, etc). For academic research, the platform provides a novel means for studying language and cognitive development at an unprecedented scale and level of detail. They will submit algorithms to the secure server which will only output anonymized aggregate statistics. Ultimately, BabyCloud aims at creating an ecosystem of third parties (public and private research labs...) gravitating around developmental data, entirely controlled by the party whose data originate from, i.e. families.

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Bilateral Grants with Industry

- Google Faculty Award - 100K€
- Facebook AI Research Grant - 350K€

# 8. Partnerships and Cooperations

## 8.1. Regional Initiatives

Collaboration with the Willow Team:

- co-advising with J. Sivic and I. Laptev of a PhD student: Ronan Riochet.
- construction of a naive physics benchmark (www.intphys.com)

## 8.2. National Initiatives

### *8.2.1. ANR*

- Transatlantic Platform "Digging into Data". Title: "Analysis of Children's Language Experiences Around the World. (ACLEW)"; (coordinating PI : M. Soderstrom; Leader of tools development and co-PI : E. Dupoux), (2017–2020. 5 countries; Total budget: 1.4M€)

## 8.3. International Initiatives

### *8.3.1. Inria International Partners*

*8.3.1.1. Informal International Partners*

- Johns Hopkins University, Baltimore, USA: S. Kudanpur, H. Hermanksy
- RIKEN Institute, Tokyo, Japan: R. Mazuka

## 8.4. International Research Visitors

### *8.4.1. Visits of International Scientists*

*8.4.1.1. Internships*

Internship of Diego Andai Castilla (partnership Inria-PUC-Inria Chile)

### 8.4.2. Visits to International Teams

*8.4.2.1. Research Stays Abroad*

- E. Dupoux Visiting Researcher at Facebook AI Research, Paris (Feb-Mar 2018)
- E. Dupoux Visiting Researcher at Google & DeepMind, London (April-July 2018)

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### 9.1.1. Scientific Events Organisation

*9.1.1.1. General Chair, Scientific Chair*

- E. Dupoux Co-Program Chair of NIPS 2018 workshop on intuitive physics, Montreal.
- E. Dupoux Co-Program chair of of the LEGRAIN Conference on Learning in Humans and Machines, Ecole Normale Supérieure, 2018 (this conference had a scientific, an industrial and a general public track)

*9.1.1.2. Member of the Organizing Committees*

- Executive committee of SIGMORPHON (Association for Computational Linguistics Special Interest Group, http://www.sigmorphon.org/).
- Executive committee of DARCLE www.darcle.org.

### 9.1.2. Scientific Events Selection

*9.1.2.1. Reviewer*

Invited editor for international conferences: Interspeech, NIPS, ACL, etc. (around 5-10 papers per conferences, 2 conferences per year)

### 9.1.3. Journal

*9.1.3.1. Member of the Editorial Boards*

Member of the editorial board of: Mathématiques et Sciences Humaines, L'Année Psychologique, Frontiers in Psychology.

*9.1.3.2. Reviewer - Reviewing Activities*

Invited Reviewer for Frontiers in Psychology, Cognitive Science, Cognition, Transactions in Acoustics Signal Processing and Language, Speech Communication, etc. (around 4 papers per year)

### 9.1.4. Invited Talks

- Nov/29/2018, E. Dupoux, Invited Department Colloquium, Linguistics, U. Maryland: Reverse Engineering Language Acquisition
- Nov/21/2018, E. Dupoux, Invited Department Colloquium, LORIA, Nancy: Developmental AI
- Oct/17/2018, E. Dupoux, Invited Seminar, Departement Physics ENS& chaire Sciences des Données: Reverse Engineering Cognitive Development
- Jul/4/2018, E. Dupoux, Invited Seminar, PRAIRIE AI Summer School: Unsupervised Speech Technology
- Nov/23/2018, E. Dupoux, Invited Seminar, GDR "Cognitive Neurosciences of Development": What AI can bring to Cognitive Development (and vice versa)
- 2018, N. Zeghidour, invited Seminar, LORIA, Nancy: learning from raw waveforms
- 2018, N. Zeghidour, invited Talk, Legrain Conference on AI and Cognition, Paris: learning from raw waveforms

### 9.1.5. Scientific Expertise

E. Dupoux is invited expert for ERC, ANR, and other granting agencies, or tenure committees (around 2 per year).

### 9.1.6. Research Administration

E. Dupoux is on the Executive committee of the Foundation Cognition, the research programme IRIS-PSL "Sciences des Données et Données des Sciences", the industrial chair Almerys (2016-) and the collective organization DARCLE (www.darcle.org).

## 9.2. Teaching - Supervision - Juries

### 9.2.1. Teaching

- Master : E. Dupoux, "Theoretical Cognitive Science: Connections and symbols", 8h, M1/M2, PSL,Paris 5, Paris France
- Master : E. Dupoux (with B. Sagot, ALMANACH, N. Zeghidour & R. Riad, COML), "Algorithms for speech and language processing", 30h, M2, (MVA), ENS Cachan, France
- Master : E. Dupoux, "Cognitive Engineering", 80h, M2, ITI-PSL, Paris France
- Doctorat : E. Dupoux, "Computational models of cognitive development", 32 h, Séminaire EHESS, Paris France

### 9.2.2. Supervision

- PhD : Julia Maria Carbajal, Separation and acquisition of two languages in early childhood: a multidisciplinary approach, Ecole Normale Supérieure, sept 21, 2018, co-advised E. Dupoux, S. Peperkamp
- PhD : Adriana Rukoz Gevara, Decoding perceptual epenthesis: Experiments and Modelling., Ecole Normale Supérieure, oct 19, 2018, co-advised E. Dupoux, S. Peperkamp
- PhD in progress : Neil Zeghidour, Learning speech features from raw signals, Feb 2015, co-advised E. Dupoux, N. Usunier (Facebook-CIFRE)
- PhD in progress : Elin Larsen, Models of word learning in infants, Sept 2017, co-advised E. Dupoux, A. Cristia– abandon
- PhD in progress : Rama Chaabouni, Language learning in artificial agents, Sept 2017, co-advised E. Dupoux, M. Baroni (Facebook-CIFRE)
- PhD in progress : Ronan Riochet, Learning models of intuitive physics, Sept 2017, co-advised E. Dupoux, I. Laptev, J. Sivic
- PhD in progress : Rachid Riad, "Speech technology for biomarkers in neurodegenerative diseases" , Sept 2018, co-advised E. Dupoux, A.-C. Bachoud-Lévi

### 9.2.3. Juries

E. Dupoux participated in the PhD Jury of Andreux Mathieu, Nov 12, ENS, 2018.

## 9.3. Popularization

E. Dupoux talked in two general public conferences on speech technologies, one organized by the Institut Carnot Cognition (La Vilette, oct, 2018), one by the Institut IA in Toulouse (oct 2018), both with around 200 participants. He gave and/or organized smaller meetings geared towards enhancing contacts between industry and research in the general area of AI and Cognition (1 day and a half of scientific meetings between PSL and Facebook, seminar-style intervention with MSR, and with the CVT Athena). He co-chaired the conference Legrain on AI and Cognition which, besides the scientific track had a general public track and an industry track, which were both attended by 100-200 attendees (see http://olivierlegrain.ens.psl.eu/ia-et-cognition.html).

N. Zeghidour did a high level presentation of AI in Vivatech on the Facebook Stand (100 000 visitors). He presented the state of the art in ASR and TTS in the BNP Paribas-PRAIRIE Summer School with participants from the IT Industry. He co-redacted a 5 pages popularization article on neural networks and deep learning in the magazine "Tangent, the mathematical adventure" for high school students, 20,000 printed copies.

# 10. Bibliography

## Major publications by the team in recent years

[1] E. DUPOUX.*Cognitive Science in the era of Artificial Intelligence: A roadmap for reverse-engineering the infant language-learner*, in "Cognition", 2018

[2] A. FOURTASSI, E. DUPOUX.*A Rudimentary Lexicon and Semantics Help Bootstrap Phoneme Acquisition*, in "Proceedings of the 18th Conference on Computational Natural Language Learning (CoNLL)", Baltimore, Maryland USA, Association for Computational Linguistics, June 2014, p. 191-200 [*DOI :* 10.3115/V1/W14-1620]

[3] A. FOURTASSI, T. SCHATZ, B. VARADARAJAN, E. DUPOUX.*Exploring the Relative Role of Bottom-up and Top-down Information in Phoneme Learning*, in "Proceedings of the 52nd Annual meeting of the ACL", Baltimore, Maryland, Association for Computational Linguistics, 2014, vol. 2, p. 1-6 [*DOI :* 10.3115/V1/P14-2001]

[4] Y. HOSHEN, R. J. WEISS, K. W. WILSON.*Speech acoustic modeling from raw multichannel waveforms*, in "Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on", IEEE, 2015, p. 4624–4628

[5] T. LINZEN, E. DUPOUX, Y. GOLDBERG.*Assessing the ability of LSTMs to learn syntax-sensitive dependencies*, in "Transactions of the Association for Computational Linguistics", 2016, vol. 4, p. 521-535

[6] T. LINZEN, E. DUPOUX, B. SPECTOR.*Quantificational features in distributional word representations*, in "Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics", 2016, p. pages 1 – 1-11 [*DOI :* 10.18653/V1/S16-2001]

[7] A. MARTIN, S. PEPERKAMP, E. DUPOUX.*Learning Phonemes with a Proto-lexicon*, in "Cognitive Science", 2013, vol. 37, p. 103-124 [*DOI :* 10.1111/J.1551-6709.2012.01267.X]

[8] S. MEHRI, K. KUMAR, I. GULRAJANI, R. KUMAR, S. JAIN, J. SOTELO, A. COURVILLE, Y. BENGIO.*SampleRNN: An unconditional end-to-end neural audio generation model*, in "arXiv preprint arXiv:1612.07837", 2016

[9] T. N. SAINATH, R. J. WEISS, A. SENIOR, K. W. WILSON, O. VINYALS.*Learning the speech front-end with raw waveform CLDNNs*, in "Sixteenth Annual Conference of the International Speech Communication Association", 2015

[10] T. SCHATZ, V. PEDDINTI, F. BACH, A. JANSEN, H. HYNEK, E. DUPOUX.*Evaluating speech features with the Minimal-Pair ABX task: Analysis of the classical MFC/PLP pipeline*, in "INTERSPEECH-2013", Lyon, France, International Speech Communication Association, 2013, p. 1781-1785

[11] R. THIOLLIÈRE, E. DUNBAR, G. SYNNAEVE, M. VERSTEEGH, E. DUPOUX.*A Hybrid Dynamic Time Warping-Deep Neural Network Architecture for Unsupervised Acoustic Modeling*, in "INTERSPEECH-2015", 2015, p. 3179-3183

[12] A. VAN DEN OORD, S. DIELEMAN, H. ZEN, K. SIMONYAN, O. VINYALS, A. GRAVES, N. KALCHBRENNER, A. SENIOR, K. KAVUKCUOGLU.*Wavenet: A generative model for raw audio*, in "CoRR abs/1609.03499", 2016

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[13] M. J. CARBAJAL.*Separation and acquisition of two languages in early childhood: A multidisciplinary approach*, Université de recherche Paris Sciences et Lettres , September 2018, https://hal.archives-ouvertes.fr/tel-01948483

[14] A. GUEVARA-RUKOZ.*Decoding perceptual vowel epenthesis: Experiments & Modelling*, Ecole Normale Supérieure (ENS), October 2018, https://hal.archives-ouvertes.fr/tel-01948548

## Articles in International Peer-Reviewed Journal

[15] E. DUPOUX.*Cognitive science in the era of artificial intelligence: A roadmap for reverse-engineering the infant language-learner*, in "Cognition", April 2018, vol. 173, p. 43 - 59, https://arxiv.org/abs/1607.08723 [*DOI :* 10.1016/J.COGNITION.2017.11.008], https://hal.archives-ouvertes.fr/hal-01888694

[16] A. GUEVARA-RUKOZ, A. CRISTIA, B. LUDUSAN, R. THIOLLIÈRE, A. MARTIN, R. MAZUKA, E. DUPOUX.*Are Words Easier to Learn From Infant- Than Adult-Directed Speech? A Quantitative Corpus-Based Investigation*, in "Cognitive Science", July 2018, vol. 42, n$^{o}$ 5, p. 1586 - 1617 [*DOI :* 10.1111/COGS.12616], https://hal.archives-ouvertes.fr/hal-01888701

[17] T. SCHATZ, F. BACH, E. DUPOUX.*Evaluating automatic speech recognition systems as quantitative models of cross-lingual phonetic category perception*, in "Journal of the Acoustical Society of America", May 2018, vol. 143, n$^{o}$ 5, p. EL372 - EL378 [*DOI :* 10.1121/1.5037615], https://hal.archives-ouvertes.fr/hal-01888735

## International Conferences with Proceedings

[18] X.-N. CAO, C. DAKHLIA, P. DEL CARMEN, M.-A. JAOUANI, M. OULD-ARBI, E. DUPOUX.*Baby Cloud, a technological platform for parents and researchers*, in "LREC 2018 - 11th edition of the Language Resources and Evaluation Conference", Miyazaki, Japan, Proceedings of LREC 2018, May 2018, https://hal.archives-ouvertes.fr/hal-01948107

[19] N. HOLZENBERGER, M. DU, J. KARADAYI, R. RIAD, E. DUPOUX.*Learning Word Embeddings: Unsupervised Methods for Fixed-size Representations of Variable-length Speech Segments*, in "Interspeech 2018", Hyderabad, India, Proceedings of Interspeech 2018, ISCA, September 2018 [*DOI :* 10.21437/INTERSPEECH.2018-2364], https://hal.archives-ouvertes.fr/hal-01888708

[20] L. ONDEL, P. GODARD, L. BESACIER, E. LARSEN, M. HASEGAWA-JOHNSON, O. SCHARENBORG, E. DUPOUX, L. BURGET, F. YVON, S. KHUDANPUR.*Bayesian Models for Unit Discovery on a Very Low Resource Language*, in "ICASSP 2018", Calgary, Alberta, Canada, Proceedings of ICASSP 2018, April 2018, https://arxiv.org/abs/1802.06053 - Accepted to ICASSP 2018, https://hal.archives-ouvertes.fr/hal-01888718

[21] R. RIAD, C. DANCETTE, J. KARADAYI, N. ZEGHIDOUR, T. SCHATZ, E. DUPOUX.*Sampling strategies in Siamese Networks for unsupervised speech representation learning*, in "Interspeech 2018", Hyderabad, India, Proceedings of Interspeech 2018, September 2018, https://arxiv.org/abs/1804.11297 - Conference paper at Interspeech 2018, https://hal.archives-ouvertes.fr/hal-01888725

[22] O. SCHARENBORG, L. BESACIER, A. BLACK, M. HASEGAWA-JOHNSON, F. METZE, G. NEUBIG, S. STUKER, P. GODARD, M. MULLER, L. ONDEL, S. PALASKAR, P. ARTHUR, F. CIANNELLA, M. DU, E. LARSEN, D. MERKX, R. RIAD, L. WANG, E. DUPOUX.*Linguistic unit discovery from multi-modal inputs in unwritten languages: Summary of the "Speaking rosetta" JSALT 2017 workshop*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Alberta, Canada, April 2018, https://hal.archives-ouvertes.fr/hal-01709578

[23] A. THUAL, C. DANCETTE, J. KARADAYI, J. BENJUMEA, E. DUPOUX.*A K-nearest neighbours approach to unsupervised spoken term discovery*, in "IEEE Spoken Language Technology SLT-2018", Athènes, Greece, Proceedings of SLT 2018, December 2018, https://hal.archives-ouvertes.fr/hal-01947953

[24] N. ZEGHIDOUR, N. USUNIER, I. KOKKINOS, T. SCHATZ, G. SYNNAEVE, E. DUPOUX.*Learning Filterbanks from Raw Speech for Phoneme Recognition*, in "ICASSP 2018 - IEEE International Conference on Acoustics, Speech and Signal Processing", Calgary, Alberta, Canada, Proceedings of ICASSP 2018, April 2018, https://arxiv.org/abs/1711.01161v2 - Accepted at ICASSP 2018, https://hal.archives-ouvertes.fr/hal-01888737

[25] N. ZEGHIDOUR, N. USUNIER, G. SYNNAEVE, R. COLLOBERT, E. DUPOUX.*End-to-End Speech Recognition From the Raw Waveform*, in "Interspeech 2018", Hyderabad, India, Proceedings of Interspeech 2018, September 2018, https://arxiv.org/abs/1806.07098 - Accepted for presentation at Interspeech 2018 [*DOI :* 10.21437/INTERSPEECH.2018-2414], https://hal.archives-ouvertes.fr/hal-01888739

### Conferences without Proceedings

[26] A. DÉFOSSEZ, N. ZEGHIDOUR, N. USUNIER, L. BOTTOU, F. BACH.*SING: Symbol-to-Instrument Neural Generator*, in "Conference on Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2018, https://arxiv.org/abs/1810.09785 , https://hal.archives-ouvertes.fr/hal-01899949

## References in notes

[27] D. A. FERRUCCI.*Introduction to "this is watson"*, in "IBM Journal of Research and Development", 2012, vol. 56, n$^{o}$ 3.4, p. 1–1

[28] K. HE, X. ZHANG, S. REN, J. SUN.*Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*, in "Proceedings of the IEEE International Conference on Computer Vision", 2015, p. 1026–1034

[29] J. HERNÁNDEZ-ORALLO, F. MARTÍNEZ-PLUMED, U. SCHMID, M. SIEBERS, D. L. DOWE.*Computer models solving intelligence test problems: Progress and implications*, in "Artificial Intelligence", 2016, vol. 230, p. 74–107

[30] B. M. LAKE, T. D. ULLMAN, J. B. TENENBAUM, S. J. GERSHMAN.*Building machines that learn and think like people*, in "arXiv preprint arXiv:1604.00289", 2016

[31] C. LU, X. TANG.*Surpassing human-level face verification performance on LFW with GaussianFace*, in "arXiv preprint arXiv:1404.3840", 2014

[32] S. T. MUELLER.*A partial implementation of the BICA cognitive decathlon using the Psychology Experiment Building Language (PEBL)*, in "International Journal of Machine Consciousness", 2010, vol. 2, n$^{o}$ 02, p. 273–288

[33] D. SILVER, A. HUANG, C. J. MADDISON, A. GUEZ, L. SIFRE, G. VAN DEN DRIESSCHE, J. SCHRITTWIESER, I. ANTONOGLOU, V. PANNEERSHELVAM, M. LANCTOT, S. DIELEMAN, D. GREWE, J. NHAM, N. KALCHBRENNER, I. SUTSKEVER, T. LILLICRAP, M. LEACH, K. KAVUKCUOGLU, T. GRAEPEL, D. HASSABIS.*Mastering the game of Go with deep neural networks and tree search*, in "Nature", 2016, vol. 529, n$^{o}$ 7587, p. 484–489

[34] I. SUTSKEVER, O. VINYALS, Q. V. LE.*Sequence to sequence learning with neural networks*, in "Advances in neural information processing systems", 2014, p. 3104–3112

[35] A. M. TURING.*Computing machinery and intelligence*, in "Mind", 1950, vol. 59, n$^{o}$ 236, p. 433–460

[36] W. XIONG, J. DROPPO, X. HUANG, F. SEIDE, M. SELTZER, A. STOLCKE, D. YU, G. ZWEIG.*Achieving human parity in conversational speech recognition*, in "arXiv preprint arXiv:1610.05256", 2016

# Team DELYS

## DistributEd aLgorithms and 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
**Paris**

THEME
**Distributed Systems and middleware**

# Table of contents

# Team DELYS

*Creation of the Team: 2018 January 01, updated into Project-Team: 2019 January 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.1.1. - Multicore, Manycore
A1.1.9. - Fault tolerant systems
A1.2.5. - Internet of things
A1.3.2. - Mobile distributed systems
A1.3.3. - Blockchain
A1.3.4. - Peer to peer
A1.3.5. - Cloud
A1.3.6. - Fog, Edge
A1.5.2. - Communicating systems
A2.6. - Infrastructure software
A2.6.1. - Operating systems
A2.6.2. - Middleware
A2.6.3. - Virtual machines
A2.6.4. - Ressource management
A3.1.8. - Big data (production, storage, transfer)
A7.1.1. - Distributed algorithms

**Other Research Topics and Application Domains:**

B6.4. - Internet of things

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Mesaac Makpangou [Inria, Researcher, HDR]
Marc Shapiro [Inria, Senior Researcher, HDR]

**Faculty Members**

Luciana Bezerra Arantes [Sorbonne Univ, Associate Professor]
Philippe Darche [Univ René Descartes, Associate Professor]
Swan Dubois [Sorbonne Univ, Associate Professor]
Jonathan Lejeune [Sorbonne Univ, Associate Professor]
Franck Petit [Sorbonne Univ, Professor, HDR]
Pierre Sens [Team leader, Sorbonne Univ, Professor, HDR]
Julien Sopena [Sorbonne Univ, Associate Professor]

**External Collaborator**

Sébastien Monnet [Univ Savoie Mont-Blanc]

**PhD Students**

Sébastien Bouchard [Inria]
Marjorie Bournat [Sorbonne Univ]
Michael Damien Carver [Magency, Sorbonne Univ]
João Paulo de Araujo [Sorbonne Univ]

Arnaud Favier [Inria, from Oct 2018]
Guillaume Fraysse [Orange Labs]
Lyes Hamidouche [Magency, until Apr 2018]
Saalik Hatia [Sorbonne Univ, since Oct 2018]
Denis Jeanneau [Sorbonne Univ]
Francis Laniel [Sorbonne Univ]
Sreeja Nair [Sorbonne Univ, since April 2018]
Jonathan Sid-Otmane [Orange Labs, Sorbonne Univ]
Alejandro Tomsic [Inria, until Apr 2018]
Ilyas Toumlilt [Sorbonne Univ]
Dimitrios Vasilas [Scality, Sorbonne Univ]
Gauthier Voron [Sorbonne Univ, until Apr 2018]

**Post-Doctoral Fellows**

Laurent Feuilloley [Sorbonne Univ]
Paolo Viotti [Sorbonne Univ, until Feb. 2018]

**Administrative Assistant**

Nelly Maloisel [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

The research of the Delys team addresses the theory and practice of distributed systems, including multicore computers, clusters, networks, peer-to-peer systems, cloud and fog computing systems, and other communicating entities such as swarms of robots. It addresses the challenges of correctly communicating, sharing information, and computing in such large-scale, highly dynamic computer systems. This includes addressing the core problems of communication, consensus and fault detection, scalability, replication and consistency of shared data, information sharing in collaborative groups, dynamic content distribution, and multi- and many-core concurrent algorithms.

Delys is a joint research team between LIP6 (UPMC/CNRS) and Inria Paris.

# 3. Research Program

## 3.1. Research rationale

DELYS addresses both theoretical and practical issues of *Computer Systems*, leveraging our dual expertise in theoretical and experimental research. Our approach is a "virtuous cycle," triggered by issues with real systems, of algorithm design which we prove correct and evaluate theoretically, and then implement and test experimentally feeding back to theory. The major challenges addressed by DELYS are the sharing of information and guaranteeing correct execution of highly-dynamic computer systems. Our research covers a large spectrum of distributed computer systems: multicore computers, mobile networks, cloud computing systems, and dynamic communicating entities. This holistic approach enables handling related problems at different levels. Among such problems we can highlight consensus, fault detection, scalability, search of information, resource allocation, replication and consistency of shared data, dynamic content distribution, and concurrent and parallel algorithms.

Two main evolutions in the Computer Systems area strongly influence our research project:

(1) Modern computer systems are **increasingly distributed**, **dynamic** and composed of multiple devices **geographically spread over heterogeneous platforms**, spanning multiple management domains. Years of research in the field are now coming to fruition, and are being used by millions of users of web systems, peer-to-peer systems, gaming and social applications, cloud computing, and now fog computing. These new uses bring new challenges, such as *adaptation to dynamically-changing conditions*, where knowledge of the system state can only be partial and incomplete.

(2) **Heterogeneous architectures and virtualisation are everywhere**. The parallelism offered by distributed clusters and *multicore* architectures is opening highly parallel computing to new application areas. To be successful, however, many issues need to be addressed. Challenges include obtaining a consistent view of shared resources, such as memory, and optimally distributing computations among heterogeneous architectures. These issues arise at a more fine-grained level than before, leading to the need for different solutions down to OS level itself.

The scientific challenges of the distributed computing systems are subject to many important features which include scalability, fault tolerance, dynamics, emergent behaviour, heterogeneity, and virtualisation at many levels. Algorithms designed for traditional distributed systems, such as resource allocation, data storage and placement, and concurrent access to shared data, need to be redefined or revisited in order to work properly under the constraints of these new environments. Sometimes, classical "*static*" problems, (*e.g.*, Election Leader, Spanning Tree Construction, ...) even need to be redefined to consider the unstable nature of the distributed system. In particular, DELYS will focus on three key challenges:

Rethinking distributed algorithms. From a theoretical point of view the key question is how to adapt the fundamental building blocks to new architectures. More specifically, how to rethink the classical algorithms to take into account the dynamics of advanced modern systems. Since a recent past, there have been several papers that propose models for dynamic systems: there is practically a different model for each setting and currently there is no unification of models. Furthermore, models often suffer of lack of realism. One of the key challenge is to identify which assumptions make sense in new distributed systems. DELYS's objectives are then (1) to identify under which realistic assumptions a given fundamental problem such as mutual exclusion, consensus or leader election can be solved and (2) to design efficient algorithms under these assumptions.

Resource management in heterogeneous systems. The key question is how to manage resources on large and heterogeneous configurations. Managing resources in such systems requires fully decentralized solutions, and to rethink the way various platforms can collaborate and interoperate with each other. In this context, data management is a key component. The fundamental issue we address in ow to efficiently and reliably share information in highly distributed environments.

Adaptation of runtimes. One of the main challenge of the OS community is how to adapt runtime supports to new architectures. With the increasingly widespread use of multicore architectures and virtualised environments, internal runtime protocols need to be revisited. Especially, memory management is crucial in OS and virtualisation technologies have highly impact on it. On one hand, the isolation property of virtualisation has severe side effects on the efficiency of memory allocation since it needs to be constantly balanced between hosted OSs. On the other hand, by hiding the physical machine to OSs, virtualisation prevents them to efficiently place their data in memory on different cores. Our research will thus focus on providing solutions to efficiently share memory between OSs without jeopardizing isolation properties.

# 4. Highlights of the Year

## 4.1. Highlights of the Year

In 2018, the DELYS team published papers at major conferences in Systems, Distributed Systems, Theoretical Computer Science, Verification, and AI:

- Scheduling under Uncertainty: A Query-based Approach. L. Arantes, E. Bampis, A. Kononov, M. Letsios, G. Lucarelli, P. Sens. IJCAI, [19].
- Byzantine Gathering in Polynomial Time. S. Bouchard, Y. Dieudonné, A. Lamani. ICALP [22].
- The Battle of the Schedulers: FreeBSD ULE vs. Linux CFS. J. Bouron, S. Chevalley, B. Lepers, W. Zwaenepoel, R. Gouicem, J. Lawall, G. Muller, J. Sopena. ATC [24].
- Distributed transactional reads: the strong, the quick, the fresh & the impossible. A. Z. Tomsic, M. Bravo, M. Shapiro. Middleware [31].
- Co-design and verification of an available file system. M. Najafzadeh, M. Shapiro, P. Eugster. VMCAI [28].

# 5. New Results

## 5.1. Distributed Algorithms for Dynamic Networks and Fault Tolerance

**Participants:** Luciana Bezerra Arantes [correspondent], Sébastien Bouchard, Marjorie Bournat, João Paulo de Araujo, Swan Dubois, Laurent Feuilloley, Denis Jeanneau, Jonathan Lejeune, Franck Petit [correspondent], Pierre Sens, Julien Sopena.

Nowadays, distributed systems are more and more heterogeneous and versatile. Computing units can join, leave or move inside a global infrastructure. These features require the implementation of *dynamic* systems, that is to say they can cope autonomously with changes in their structure in terms of physical facilities and software. It therefore becomes necessary to define, develop, and validate distributed algorithms able to managed such dynamic and large scale systems, for instance mobile *ad hoc* networks, (mobile) sensor networks, P2P systems, Cloud environments, robot networks, to quote only a few.

The fact that computing units may leave, join, or move may result of an intentional behavior or not. In the latter case, the system may be subject to disruptions due to component faults that can be permanent, transient, exogenous, evil-minded, etc. It is therefore crucial to come up with solutions tolerating some types of faults.

In 2018, we obtained the following results.

### *5.1.1. Scheduling in uncertain environments*

In [19], we consider scheduling with faults/errors and we introduce a new non-probabilistic model with explorable (query-able) uncertainty. Each unit-time error is characterized by an uncertainty area during which the error will occur, and it is possible to learn the exact slot at which it will appear by issuing a query operation of unit cost. We study two problems: (i) the error-query scheduling problem, whose aim is to reveal enough error-free slots with the minimum number of queries, and (ii) the lexicographic error-query scheduling problem where we seek the earliest error-free slots with the minimum number of queries. We consider both the off-line and the on-line versions of the above problems. In the former, the whole instance and its characteristics are known in advance and we give a polynomial-time algorithm for the error-query scheduling problem. In the latter, the adversary has the power to decide, in an on-line way, the time-slot of appearance for each error. We propose then both lower bounds and algorithms whose competitive ratios asymptotically match these lower bounds.

### *5.1.2. Failure detectors in dynamic systems*

The failure detector abstraction was introduced as a way to circumvent the impossibility of solving consensus in asynchronous systems prone to crash failures. A failure detector is a local oracle that provides processes in the system with unreliable information on process failures. But a failure detector that is sufficient to solve a given problem in a static system is not necessarily sufficient to solve the same problem in a dynamic system. In [37], we adapt an existing failure detector for mutual exclusion and prove that it is the weakest failure detector to solve mutual exclusion in dynamic systems, which means that it is weaker than any other failure detector capable of solving mutual exclusion.

We also propose in [15] a new failure detector, called the Impact failure detector (FD), that expresses the confidence with regard to the system as a whole. Similarly to a reputation approach, it is possible to indicate the relative importance of each process of the system, while a threshold offers a degree of flexibility for failures and false suspicions. Performance evaluation results, based on real PlanetLab traces, confirm the degree of flexible of the failure detector.

### *5.1.3. Causal information dissemination*

A causal broadcast ensures that messages are delivered to all nodes (processes) preserving causal relation of the messages. In [33], we propose a new causal broadcast protocol for distributed systems whose nodes are logically organized in a virtual hypercube-like topology called VCube. Messages are broadcast by dynamically building spanning trees rooted in the message's source node. By using multiple trees, the contention bottleneck problem of a single root spanning tree approach is avoided. Furthermore , different trees can intersect at some node. Hence, by taking advantage of both the out-of-order reception of causally related messages at a node and these paths intersections, a node can delay to one or more of its children in the tree. Experimental evaluation conducted on top of PeerSim simulator confirms the communication effectiveness of our causal broadcast protocol in terms of latency and message traffic reduction

### *5.1.4. Graceful Degradation*

Gracefully degrading algorithms was introduced by Biely *et al.*i. Such algorithms offer the desirable properties to circumvent impossibility results in dynamic systems by adapting themselves to the dynamics. Indeed, such a algorithms solve a given problem under some dynamics and, moreover, guarantees that a weaker (but related) problem is solved under a higher dynamics under which the original problem is impossible to solve. The underlying intuition is to solve the problem whenever possible but to provide some kind of quality of service if the dynamics become (unpredictably) higher.

In [36], we apply for the first time this approach to robot networks. We focus on the fundamental problem of gathering a squad of autonomous robots on an unknown location of a dynamic ring. In this goal, we introduce a set of weaker variants of this problem. Motivated by a set of impossibility results related to the dynamics of the ring, we propose a gracefully degrading gathering algorithm.

### *5.1.5. Unreliable Hints*

In [23], we address the question of a mobile agent deterministically searching for a target in the Euclidean plane. We assume that the mobile agent is equipped with a compass and a measure of length has to find an inert treasure in the Euclidean plane. Both the agent and the treasure are modeled as points. In the beginning, the agent is at a distance at most $D > 0$ from the treasure, but knows neither the distance nor any bound on it. Finding the treasure means getting at distance at most 1 from it. The agent makes a series of moves. Each of them consists in moving straight in a chosen direction at a chosen distance. In the beginning and after each move the agent gets a hint consisting of a positive angle smaller than $2\pi$ whose vertex is at the current position of the agent and within which the treasure is contained. We investigate the problem of how these hints permit the agent to lower the cost of finding the treasure, using a deterministic algorithm, where the cost is the worst-case total length of the agent's trajectory. It is well known that without any hint the optimal (worst case) cost is $\Theta(D^2)$. We show that if all angles given as hints are at most $\pi$, then the cost can be lowered to $O(D)$, which is optimal. If all angles are at most $\beta$, where $\beta < 2\pi$ is a constant unknown to the agent, then the cost is at most $O(D^2 - \epsilon)$, for some $\epsilon > 0$. For both these positive results we present deterministic algorithms achieving the above costs. Finally, if angles given as hints can be arbitrary, smaller than $2\pi$, then we show that cost $\Theta(D^2)$ cannot be beaten.

### *5.1.6. Gathering of Mobile Agents*

Gathering a group of mobile agents is a fundamental task in the field of distributed and mobile systems. It consists of bringing agents that initially start from different positions to meet all together in finite time. In the case when there are only two agents, the gathering problem is often referred to as the rendezvous problem.

In [14] and [22], we consider these tasks from a deterministic point of view in networks modeled as undirected and anonymous graphs. An adversary chooses the initial nodes of the agents (the number of agents may be larger than the number of nodes) and assigns a different positive integer (called label) to each of them. Initially, each agent knows its label as well as some global knowledge shared by all the agents. The agents can communicate with each other only when located at the same node.

This task has been considered in the literature under two alternative scenarios: weak and strong. Under the weak scenario, agents may meet either at a node or inside an edge. Under the strong scenario, they have to meet at a node, and they do not even notice meetings inside an edge. Gathering and rendezvous algorithms under the strong scenario are known for synchronous agents. For asynchronous agents, gathering and rendezvous under the strong scenario are impossible even in the two-node graph, and hence only algorithms under the weak scenario were constructed.

In [14] we show that rendezvous under the strong scenario is possible for agents with asynchrony restricted in the following way: agents have the same measure of time but the adversary can impose, for each agent and each edge, the speed of traversing this edge by this agent. The speeds may be different for different edges and different agents but all traversals of a given edge by a given agent have to be at the same imposed speed. We construct a deterministic rendezvous algorithm for such agents, working in time polynomial in the size of the graph, in the length of the smaller label, and in the largest edge traversal time.

Gathering mobile agents can be made drastically more difficult to achieve when some agents are subject to faults, especially the Byzantine ones that are known as being the worst faults to handle. Byzantine means that the agent is subject to unpredictable and arbitrary faults. For instance, such an agent may choose to never stop or to never move. In [22] we study the task of Byzantine gathering among synchronous agents under the strong scenario: despite the presence of $f$ Byzantine agents, all the other (correct) agents have to meet at the same node. In this respect, assuming that the agents are in a *strong team* i.e., a team in which the number of correct agents is at least some prescribed value that is quadratic in $f$, we show an algorithm that solves Byzantine gathering with all strong teams in all graphs of size at most $n$, for any integers $n$ and $f$, in a time polynomial in $n$ and the length $|l_{min}|$ of the binary representation of the smallest label of a good agent. The algorithm works using a global knowledge of size $\mathcal{O}(\log\log\log n)$, which we prove to be of optimal order of magnitude in our context to reach a time complexity that is polynomial in $n$ and $|l_{min}|$.

### 5.1.7. Self-Stabilizing Minimum Diameter Spanning Tree

In [13], we present a self-stabilizing algorithm for the minimum diameter spanning tree construction problem in the state model. Our protocol has the following attractive features. It is the first algorithm for this problem that operates under the *unfair and distributed* adversary (or *daemon*). In other words, no restriction is made on the asynchronous behavior of the system. Second, our algorithm needs only $O(\log n)$ bits of memory per process (where $n$ is the number of processes), that improves the previous result by a factor $n$. These features are not achieved to the detriment of the convergence time, which stays polynomial.

## 5.2. Large-scale data distribution

**Participants:** Saalik Hatia, Mesaac Makpangou, Sébastien Monnet, Sreeja Nair, Jonathan Sid-Otmane, Pierre Sens, Marc Shapiro, Alejandro Tomsic, Ilyas Toumlilt, Dimitrios Vasilas, Paolo Viotti.

### 5.2.1. Impossibility results for distributed transactional reads

We study the costs and trade-offs of providing transactional consistent reads in a distributed storage system. We identify the following dimensions: read consistency, read delay (latency), and data freshness. We show that there is a three-way trade-off between them, which can be summarised as follows: (i) it is not possible to ensure at the same time order-preserving (e.g., causally-consistent) or atomic reads, Minimal Delay, and maximal freshness; thus, reading data that is the most fresh without delay is possible only in a weakly-isolated mode; (ii) to ensure atomic or order-preserving reads at Minimal Delay imposes to read data from the past (not fresh); (iii) however, order-preserving minimal-delay reads can be fresher than atomic; (iv) reading atomic or order-preserving data at maximal freshness may block reads or writes indefinitely. Our impossibility results

hold independently of other features of the database, such as update semantics (totally ordered or not) or data model (structured or unstructured). Guided by these results, we modify an existing protocol to ensure minimal-delay reads (at the cost of freshness) under atomic-visibility and causally-consistent semantics. Our experimental evaluation supports the theoretical results.

This work was published at Middleware 2018 [31].

### *5.2.2. Co-design and verification of an available file system*

Distributed file systems play a vital role in large-scale enterprise services. However, the designer of a distributed file system faces a vexing choice between strong consistency and asynchronous replication. The former supports a standard sequential model by synchronising operations, but is slow and fragile. The latter is highly available and responsive, but exposes users to concurrency anomalies. We describe a rigorous and general approach to navigating this trade-off by leveraging static verification tools that allow to verify different file system designs. We show that common file system operations can run concurrently without synchronisation, while still retaining a semantics reasonably similar to Posix hierarchical structure. The one exception is the "move" operation, for which we prove that, unless synchronised, it will have an anomalous behaviour.

This work was published at VMCAI 2018 [28].

## 5.3. Resources management in system software

**Participants:** Michael Damien Carver, Jonathan Lejeune, Pierre Sens, Julien Sopena [correspondent], Gauthier Voron, Francis Laniel.

### *5.3.1. Multicore schedulers*

In collaboration with WHISPER team, we have contributed to an analysis of the impact on application performance of the design and implementation choices made in two widely used open-source schedulers: ULE, the default FreeBSD scheduler, and CFS, the default Linux scheduler. In a paper published at USENIX ATC'18 [24], we compare ULE and CFS in otherwise identical circumstances. This work involves porting ULE to Linux, and using it to schedule all threads that are normally scheduled by CFS. We compare the performance of a large suite of applications on the modified kernel running ULE and on the standard Linux kernel running CFS. The observed performance differences are solely the result of scheduling decisions, and do not reflect differences in other subsystems between FreeBSD and Linux. We found that there is no overall winner. On many workloads the two schedulers perform similarly, but for some workloads there are significant and even surprising differences. ULE may cause starvation, even when executing a single application with identical threads, but this starvation may actually lead to better application performance for some workloads. The more complex load balancing mechanism of CFS reacts more quickly to workload changes, but ULE achieves better load balance in the long run.

# 6. Bilateral Contracts and Grants with Industry

## 6.1. Bilateral Contracts with Industry

DELYS has a CIFRE contract with Scality SA:

- Dimitrios Vasilas is advised by Marc Shapiro and Brad King. He works on secondary indexing in large-scale storage systems under weak consistency.

DELYS has two CIFRE contracts with Magency SA:

- Damien Carver is advised by Julien Sopena and Sébatien Monnet. He works on designing kernel-level mechanisms that automatically give more memory to the most active containers.
- Lyes Hamidouche is advised by Pierre Sens and Sébatien Monnet. He works on efficient data dissemination among a large number of mobile devices. He defended his thesis in April 2018.

DELYS has two contracts with Orange within the I/O Lab joint laboratory:

- Guillaume Fraysse is advised by Jonathan Lejeune, Julien Sopena, and Pierre Sens. He works on distributed resources allocation in virtual network environments.
- Jonathan Sid-Otmane is advised by Marc Shapiro. He studies the applications of distributed databases to the needs of the telco industry in the context of 5G.

# 7. Partnerships and Cooperations

## 7.1. National Initiatives

### *7.1.1. ANR*

*7.1.1.1. ESTATE - (2016–2020)*

Members: LIP6 (DELYS, project leader), LaBRI (Univ. de Bordeaux); Verimag (Univ. de Grenoble).

Funding: ESTATE is funded by ANR (PRC) for a total of about 544 000 euros, of which 233 376 euros for DELYS.

Objectives: The core of ESTATE consists in laying the foundations of a new algorithmic framework for enabling Autonomic Computing in distributed and highly dynamic systems and networks. We plan to design a model that includes the minimal algorithmic basis allowing the emergence of dynamic distributed systems with self-* capabilities, *e.g.*, self-organization, self-healing, self-configuration, self-management, self-optimization, self-adaptiveness, or self-repair. In order to do this, we consider three main research streams:

$(i)$ building the theoretical foundations of autonomic computing in dynamic systems, $(ii)$ enhancing the safety in some cases by establishing the minimum requirements in terms of amount or type of dynamics to allow some strong safety guarantees, $(iii)$ providing additional formal guarantees by proposing a general framework based on the Coq proof assistant to (semi-)automatically construct certified proofs.

The coordinator of ESTATE is Franck Petit.

*7.1.1.2. RainbowFS - (2016–2020)*

Members: LIP6 (DELYS, project leader), Scality SA, CNRS-LIG, Télécom Sud-Paris, Université Savoie-Mont-Blanc.

Funding: is funded by ANR (PRC) for a total of 919 534 euros, of which 359 554 euros for DELYS.

Objectives: RainbowFS proposes a "just-right" approach to storage and consistency, for developing distributed, cloud-scale applications. Existing approaches shoehorn the application design to some predefined consistency model, but no single model is appropriate for all uses. Instead, we propose tools to co-design the application and its consistency protocol. Our approach reconciles the conflicting requirements of availability and performance vs. safety: common-case operations are designed to be asynchronous; synchronisation is used only when strictly necessary to satisfy the application's integrity invariants. Furthermore, we deconstruct classical consistency models into orthogonal primitives that the developer can compose efficiently, and provide a number of tools for quick, efficient and correct cloud-scale deployment and execution. Using this methodology, we will develop an entreprise-grade, highly-scalable file system, exploring the rainbow of possible semantics, and we demonstrate it in a massive experiment.

The coordinator of RainbowFS is Marc Shapiro.

### *7.1.2. LABEX*

*7.1.2.1. SMART - (2012–2019)*

Members: ISIR (Sorbonne Univ./CNRS), LIP6 (Sorbonne Univ./CNRS), LIB (Sorbonne Univ./INSERM), LJLL (Sorbonne Univ./CNRS), LTCI (Institut Mines-Télécom/CNRS), CHArt-LUTIN (Univ. Paris 8/EPHE), L2E (Sorbonne Univ.), STMS (IRCAM/CNRS).

Funding: Sorbonne Universités, ANR.

Description: The SMART Labex project aims globally to enhancing the quality of life in our digital societies by building the foundational bases for facilitating the inclusion of intelligent artifacts in our daily life for service and assistance. The project addresses underlying scientific questions raised by the development of Human-centered digital systems and artifacts in a comprehensive way. The research program is organized along five axes and DELYS is responsible of the axe "Autonomic Distributed Environments for Mobility."

The project involves a PhD grant of 100 000 euros over 3 years.

## 7.2. European Initiatives

### *7.2.1. FP7 & H2020 Projects*

*7.2.1.1. LightKone*

Title: Lightweight Computation for Networks at the Edge

Programm: H2020-ICT-2016-2017

Duration: January 2017 - December 2019

Coordinator: Université Catholique de Louvain

Partners:

Université Catholique de Louvain (Belgium)

Technische Universitaet Kaiserslautern (Germany)

INESC TEC - Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciencia (Portugal)

Faculdade de Ciencias E Tecnologiada Universidade Nova de Lisboa (Portugal)

Universitat Politecnica De Catalunya (Spain)

Scality (France)

Gluk Advice B.V. (Netherlands)

Inria contact: Marc Shapiro

The goal of LightKone is to develop a scientifically sound and industrially validated model for doing general-purpose computation on edge networks. An edge network consists of a large set of heterogeneous, loosely coupled computing nodes situated at the logical extreme of a network. Common examples are networks of Internet of Things, mobile devices, personal computers, and points of presence including Mobile Edge Computing. Internet applications are increasingly running on edge networks, to reduce latency, increase scalability, resilience, and security, and permit local decision making. However, today's state of the art, the gossip and peer-to-peer models, give no solution for defining general-purpose computations on edge networks, i.e., computation with shared mutable state. LightKone will solve this problem by combining two recent advances in distributed computing, namely synchronisation-free programming and hybrid gossip algorithms, both of which are successfully used separately in industry. Together, they are a natural combination for edge computing. We will cover edge networks both with and without data center nodes, and applications focused on collaboration, computation, and both. Project results will be new programming models and algorithms that advance scientific understanding, implemented in new industrial applications and a startup company, and evaluated in large-scale realistic settings.

## 7.3. International Initiatives

### *7.3.1. Participation in Other International Programs*

*7.3.1.1. STIC Amsud*

Title: PaDMetBio - Parallel and Distributed Metaheuristics for Structural Bioinformatics

International Partners (Institution - Laboratory - Researcher):

Universidade Federal do Rio Grande do Sul (Brazil)- Mãrcio Dorn

Universidad Nacional de San Luis (Argentina) - Verõnica Gil-Costa

Universidad de Santiago de Chile (Chile) - Mario Inostroza-Ponta

Duration: 2017 - 2018

Start year: 2017

Structural bioinformatics deals with problems where the rules that govern the biochemical processes and relations are partially known which makes hard to design efficient computational strategies for these problems. There is a wide range of unanswered questions, which cannot be answered neither by experiments nor by classical modeling and simulation approaches. Specifically, there are several problems that still do not have a computational method that can guarantee a minimum quality of solution. Two of the main challenging problems in Structural Bioinformatics are (1) the three-dimensional (3D) protein structure prediction problem (PSP) and (2) the molecular docking problem for drug design. Predicting the folded structure of a protein only from its amino acid sequence is a challenging problem in mathematical optimization. The challenge arises due to the combinatorial explosion of plausible shapes, where a long amino acid chain ends up in one out of a vast number of 3D conformations. The problem becomes harder when we have proteins with complex topologies, in this case, their predictions may be only possible with significant increases in high-performance computing power. In the case of the molecular docking problem for drug design, we need to predict the preferred orientation of a small drug candidate against a protein molecule. With the increasing availability of molecular biological structures, smarter docking approaches have become necessary. These two problems are classified as NP-Complete or NP-Hard, so there is no current computational approach that can guarantee the best solution for them in a polynomial time. Because of the above, there is the need to build smarter approaches that can deliver good solutions to the problem. In this project, we plan to explore a collaborative work for the design and implementation of population based metaheuristics, like genetic and memetic algorithms. Metaheuristics are one of the most common and powerful techniques used in this case. The main goal of this project is to gather the expertise and current work of researchers in the areas of structural bioinformatics, metaheuristics and parallel and distributed computing, in order to build novel and high quality solutions for these hot research area.

*7.3.1.2. Capes-Cofecub*

Title: CHOOSING - Cooperation on Hybrid cOmputing clOuds for energy SavING

French Partners: Paris XI (LRI), Regal, LIG, SUPELEC

International Partners (Institution - Laboratory - Researcher):

Universidade de São Paulo - Instituto de Matemática e Estatística - Brazil, Unicamp - Instituto de Computação - Brazil

Duration: 2014–2018

The cloud computing is an important factor for environmentally sustainable development. If, in the one hand, the increasing demand of users drive the creation of large datacenters, in the other hand, cloud computing's "multitenancy" trait allows the reduction of physical hardware and, therefore, the saving of energy. Thus, it is imperative to optimize the energy consumption corresponding to the datacenter's activities. Three elements are crucial on energy consumption of a cloud platform: computation (processing), storage and network infrastructure. Therefore, the aim of this project is

to provide different techniques to reduce energy consumption regarding these three elements. Our work mainly focuses on energy saving aspects based on virtualization, i.e., pursuing the idea of the intensive migration of classical storage/processing systems to virtual ones. We will study how different organizations (whose resources are combined as hybrid clouds) can cooperate with each other in order to minimize the energy consumption without the detriment of client requirements or quality of service. Then, we intend to propose efficient algorithmic solutions and design new coordination mechanisms that incentive cloud providers to collaborate.

*7.3.1.3. Spanish research ministry project*

Title: BFT-DYNASTIE - Byzantine Fault Tolerance: Dynamic Adaptive Services for Partitionable Systems

French Partners: Labri, Irisa, LIP6

International Partners (Institution - Laboratory - Researcher):

University of the Basque Country UPV - Spain, EPFL - LSD - Switzerland, Friedrich-Alexander-Universitat Erlangen-Nurenberg - Deutschland, University of Sydney - Australia

Duration: 2017–2019

The project BFT-DYNASTIE is aimed at extending the model based on the alternation of periods of stable and unstable behavior to all aspects of fault-tolerant distributed systems, including synchrony models, process and communication channel failure models, system membership, node mobility, and network partitioning. The two main and new challenges of this project are: the consideration of the most general and complex to address failure model, known as Byzantine, arbitrary or malicious, which requires qualified majorities and the use of techniques form the security area; and the operation of the system in partitioned mode, which requires adequate reconciliation mechanisms when two partitions merge.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. General Chair, Scientific Chair*

Marc Shapiro, Organiser of Dagstuhl Workshop on “Data Consistency in Distributed Systems: Algorithms, Programs, and Databases” (19-0117), February 2018.

Swan Dubois, PC Track Chair (track A: Theoretical and Practical Aspects of Stabilizing Systems) of the 20th International Symposium on Stabilization, Safety, and Security of Distributed Systems (SSS 2018).

*8.1.1.2. Member of the Organizing Committees*

- Marc Shapiro, Steering Committee of Workshop on Principles and Practice of Consistency for distributed Data (PaPoC).
- Franck Petit, Steering Committee of the *International Symposium on Stabilization, Safety, and Security of Distributed Systems* (SSS).
- Pierre Sens, since 2014: Member of Steering Committee of International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD).

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Member of Conference Program Committees*

Pierre Sens, 28th International Symposium on Software Reliability Engineering (ISSRE 2018), 17th IEEE International Symposium on Network Computing and Applications (NCA 2018), 37th IEEE International Symposium on Reliable Distributed Systems (SRDS 2018), 32rd International Symposium on DIStributed Computing (DISC 2018).

Marc Shapiro, Program Committee of Int. Conf. on Middleware (Middleware 2018).

Marc Shapiro, Program Committee of European Conference on Computer Systems (EuroSys 2019).

Marc Shapiro, Program Committee of Symposium on Principles of Distributed Computing (PODC 2019).

Dimitri Vasilas, Shadow Program Committee of European Conference on Computer Systems (EuroSys 2019).

Franck Petit, 20th Workshop on Advances in Parallel and Distributed Computational Models (APDCM 2018), 24th Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS 2018), 20th Rencontres Francophones sur les Aspects Algorithmiques de Télécommunications (ALGOTEL 2018). Tutorial chairman of Latin-American Symposium on Dependable Computing (LADC 2018).

Luciana Arantes, 17th IEEE International Symposium on Network Computing and Applications (NCA 2018), 14th European Dependable Computing Conference (EDCC 2018), Latin-American Symposium on Dependable Computing (LADC 2018), Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS 2018) .

Swan Dubois, PC member of the ACM Symposium on Principles of Distributed Computing (PODC 2018).

### *8.1.3. Journal*

*8.1.3.1. Member of Editorial Boards*

Pierre Sens, Associate editor of International Journal of High Performance Computing and Networking (IJHPCN)

Franck Petit, Special Issue on Stabilization, Safety, and Security, Journal on Theory of Computing Systems (ToCS).

Marc Shapiro, Associate Editor for Letters of the IEEE Computer Society (LOCS).

*8.1.3.2. Reviewer, Reviewing Activities*

Pierre Sens, reviewer Journal of Parallel and Distributed Systems (JPDC), and IEEE Transactions on Parallel and Distributed Systems (TPDS).

Franck Petit, reviewer Journal of the ACM (JACM), Journal of Parallel and Distributed Systems (JPDC), and the International Colloquium on Structural Information and Communication Complexity (SIROCCO 2018).

Luciana Arantes, reviewer Journal of Parallel and Distributed Systems (JPDC), and IEEE Transactions on Parallel and Distributed Systems (TPDS).

Swan Dubois, reviewer Theoretical Computer Science (TCS), ACM Transactions on Computer Systems (TOCS), and The International Symposium on DiStributed Computing (DISC 2018).

### *8.1.4. Invited Talks*

Pierre Sens, *Fault tolerance in dynamic distributed systems*. Invited keynote speaker, "Insights for the Future of Computing", LIG, Grenoble, Avril 2018.

Franck Petit, *Robustness: a New Form of Heredity Motivated by Dynamic Networks*, Invited speaker, 9th Workshop on GRAph Searching, Theory & Applications (GRASTA 2018), Berlin, Germany, Sept. 2018.

Marc Shapiro, *Just-Right Consistency: As available as possible, As consistent as necessary, Correct by design*. Invited speaker, Verification of Distributed Systems workshop, Essaouira, Morocco, May 2018.

Marc Shapiro, *Just-Right Consistency: As available as possible, As consistent as necessary, Correct by design*. Keynote presentation, DotScale, the European Tech Conference on Scalability, Distributed Systems & DevOps, Aubervillers, June 2018.

Marc Shapiro, *Just-Right Consistency*. Keynote presentation, Conf on Advances and Computing and Communication Engineering, Paris, June 2018.

Marc Shapiro, *Just-Right Consistency: As available as possible, Synchronous when necessary, Correct by design*. Invited Keynote talk, I/O Labs annual workshop, Châtillon, France, Oct. 2018.

Marc Shapiro, *Life after consistency*. Workshop of the EuroSys Program Committee, Dec. 2018.

### *8.1.5. Scientific Expertise*

Pierre Sens, Project in Indo-French Centre for the Promotion of Advanded Research

Marc Shapiro, member of Panel PE6 (Computer Science) of European Research Council Starting Grants 2018.

### *8.1.6. Research Administration*

Franck Petit, since 2014: deputy director of the LIP6 laboratory

Pierre Sens, since 2016: Member of Section 6 of the national committee for scientific research CoNRS

Pierre Sens, since 2012: Member of the Executive Committee of Labex SMART, Co-Chair (with F. Petit) of Track 4, Autonomic Distributed Environments for Mobility.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Julien Sopena is Member of “Directoire des formations et de l’insertion professionnelle” of Sorbonne Université, France

Master: Julien Sopena is responsible of Computer Science Master’s degree in Distributed systems and applications (in French, SAR), Sorbonne Universités, France

Master: Luciana Arantes, Swan Dubois, Jonathan Lejeune, Franck Petit, Pierre Sens, Julien Sopena, Advanced distributed algorithms, M2, Sorbonne Université, France

Master: Jonathan Lejeune, Designing Large-Scale Distributed Applications, M2, Sorbonne Université, France

Master: Maxime Lorrillere, Julien Sopena, Linux Kernel Programming, M1, Sorbonne Université, France

Master: Luciana Arantes, Swan Dubois, Jonathan Lejeune, Pierre Sens, Julien Sopena, Operating systems kernel, M1, Sorbonne Université, France

Master: Luciana Arantes, Swan Dubois, Franck Petit, Distributed Algorithms, M1, Sorbonne Université, France

Master: Jonathan Lejeune, Julien Sopena, Client-server distributed systems, M1, Sorbonne Université, France.

Master: Julien Sopena, Marc Shapiro, Ilyas Toumlilt, Francis Laniel. Kernels and virtual machines (*Noyaux et machines virtuelles*, NMV), M2, Sorbonne Université, France.

Licence: Pierre Sens, Luciana Arantes, Julien Sopena, Principles of operating systems, L3, UPMC Sorbonne Université, France

Licence: Swan Dubois, Initiation to operating systems, L3, Sorbonne Université, France

Licence: Swan Dubois, Multi-threaded Programming, L3, Sorbonne Université, France

Licence: Jonathan Lejeune, Oriented-Object Programming, L3, Sorbonne Université, France

Licence: Franck Petit, Advanced C Programming, L2, Sorbonne Université, France

Licence: Swan Dubois, Julien Sopena, Introduction to operating systems, L2, Sorbonne Université, France

Licence: Mesaac Makpangou, C Programming Language, 27 h, L2, Sorbonne Université, France

Ingénieur 4ème année : Marc Shapiro, Introduction aux systèmes d'exploitation, 26 h, M1, Polytech Sorbonne Université, France.

Licence : Philippe Darche (coordinator), Architecture of Internet of Things (IoT), 2 × 32h, L3, Institut Universitaire Technologique (IUT) Paris Descartes, France.

Engineering School: Philippe Darche (coordinator), Solid-State Memories, 4th year, ESIEE, France.

DUT: Philippe Darche (coordinator), Introduction to Computer Systems - Data representation, 60h, Institut Universitaire Technologique (IUT) Paris Descartes, France.

DUT: Philippe Darche (coordinator), Computer Architecture, 32h, Institut Universitaire Technologique (IUT) Paris Descartes, France.

DUT: Philippe Darche (coordinator), Computer Systems Programming, 80h, Institut Universitaire Technologique (IUT) Paris Descartes, France.

### *8.2.2. Supervision*

PhD: Florent Coriat, "Géolocalisation et routage en situation de crise", Dec. 2018, Sorbonne Univ., Anne Fladenmuller (NPA-LIP6) and Luciana Arantes.

PhD: Lyes Hamidouche, "Data replication and data sharing in mobile networks", Sorbonne Univ., CIFRE, Apr. 2018, Sébastien Monnet, Pierre Sens, Dimitri Refauvelet (Magency).

PhD: Denis Jeanneau, "Failure detectors in Dynamic Systems," Sorbonne Université, Dec. 2018, Luciana Arantes, Pierre Sens.

PhD: B Ngom. "FreeCore: a system for indexing document summaries on a Distributed Hash Table (DHT)," Sorbonne Université, Jul. 2018, Mesaac Makpangou.

PhD: Alejandro Z. Tomsic, Sorbonne Univ., defended Apr. 2018, Marc Shapiro. "Computing over widely-replicated data in a hybrid cloud."

PhD: Gauthier Voron, "Big-Os : un OS pour les grands volumes de données,", Sorbonne Univ., Mar. 2018, Gaël Thomas, Pierre Sens.

PhD in progress: João Paulo de Araujo, "L'exécution efficace d'algorithmes distribués dans les réseaux véhiculaires", funded by CNPq (Brésil), since Nov.2015, Pierre Sens and Luciana Arantes.

PhD in progress: Sébastien Bouchard, "Gathering with faulty robots", Sorbonne Univ., since Oct. 2016, Swan Dubois, Franck Petit, Yoann Dieudonné (University of Picardy Jules Verne)

PhD in progress: Marjorie Bournat, "Speculation and Graceful Degradability for Robots in Highly Dynamic Environments", Sorbonne Univ., since Sep. 2015, Swan Dubois, Franck Petit, Yoann Dieudonné (University of Picardy Jules Verne)

PhD in progress: Damien Carver, "HACHE : HorizontAl Cache cHorEgraphy - Toward automatic resizing of shared I/O caches.", Sorbonne Univ., CIFRE, since Jan. 2015, Sébastien Monnet, Pierre Sens, Julien Sopena, Dimitri Refauvelet (Magency).

PhD in progress: Arnaud Favier, "Algorithmes de coordination répartis dans des réseaux dynamiques", Sorbonne Univ., since Sep. 2018, Pierre Sens and Luciana Arantes.

PhD in progress: Saalik Hatia, "Efficient management of memory and storage for CRDTs," Sorbonne Univ., since Oct. 2018. Advised by Marc Shapiro.

CIFRE PhD in progress: Guillaume Fraysse, Orange Lab - Inria, “Ubiquitous Resouces for Service Availability.” Since Jul. 2017, advised by Pierre Sens, Imen Grida Ben Yahia (Orange-Lab) , Jonathan Lejeune, Julien Sopena.

PhD in progress: Francis Laniel, Sorbonne Univ., since Sept. 2017. Advised by Marc Shapiro, Julien Sopena, Jonathan Lejeune. “Vers une utilisation efficace de la mémoire non volatile pour économiser l’énergie.”

PhD in progress: Sreeja Nair, Sorbonne Univ., since April 2018. “Just-Right Consistency for massive gee-replicated storage.” Advised by Marc Shapiro.

CIFRE PhD in progress: Jonathan Sid-Otmane. “Étude des critères de distribution et de l’usage d’une base de données distribuée pour un OS Telco.” Advised by Marc Shapiro, with Sofiane Imadali and Frédéric Martelli, Orange Labs.

PhD in progress: Ilyas Toumlilt, Sorbonne Univ. “Bridging the CAP gap, all the way to the edge.” Advised by Marc Shapiro.

CIFRE PhD in progress: Dimitrios Vasilas, Sorbonne Univ., “Indexing in large-scale storage systems.” Advised by Marc Shapiro, with Brad King, Scality.

### *8.2.3. Juries*

Franck Petit was the reviewer of:

- Alessia Milani, HDR, LaBRI, Bordeaux
- Laurent Feuilloley, PhD, IRIF, Paris

Franck Petit was Chair of

- Florent Coriat, PhD, LIP6, Paris

Pierre Sens was the reviewer of:

- Matthieu Roy, HDR, LAAS, Toulouse
- Thouraya Louati, PhD, Sfax Univ., Tunisie
- Yacine Taleb, PhD, IRSIA, Rennes
- Ye Xia, PhD, LIG, Univ. Grenoble

Pierre Sens was Chair of

- Soraya Zertal, HDR, UVSQ, Versailles
- Nikolaos Georgantas, HDR, Sorbonne Univ. - Inria, Paris
- Alejandro Tomsic, PhD, LIP6, Paris
- Veronica Quintuna Rodriguez, PhD, Sorbonne Univ. - Inria, Paris

Marc Shapiro was a member of the following committees:

- Aurélie Hurault, HdR, ENSEEIHT, Toulouse.
- Soraya Zertal, Comité de suivi doctoral, UVSQ, Versailles.

## 8.3. Popularization

Jonathan Lejeune and Julien Sopena animated an activity during the Science Festival 2018 at Sorbonne Univ.

# 9. Bibliography

## Major publications by the team in recent years

[1] V. BALEGAS, N. PREGUIÇA, R. RODRIGUES, S. DUARTE, C. FERREIRA, M. NAJAFZADEH, M. SHAPIRO.*Putting Consistency back into Eventual Consistency*, in "Euro. Conf. on Comp. Sys. (EuroSys)", Bordeaux, France, April 2015, p. 6:1–6:16, https://doi.org/10.1145/2741948.2741972

[2] L. GIDRA, G. THOMAS, J. SOPENA, M. SHAPIRO, N. NGUYEN.*NumaGiC: a garbage collector for big data on big NUMA machines*, in "Int. Conf. on Archi. Support for Prog. Lang. and Systems (ASPLOS)", Istanbul, Turkey, Assoc. for Computing Machinery, March 2015, p. 661–673, http://dx.doi.org/10.1145/2694344.2694361

[3] A. GOTSMAN, H. YANG, C. FERREIRA, M. NAJAFZADEH, M. SHAPIRO.*'Cause I'm Strong Enough: Reasoning about Consistency Choices in Distributed Systems*, in "Symp. on Principles of Prog. Lang. (POPL)", St. Petersburg, FL, USA, 2016, p. 371–384, http://dx.doi.org/10.1145/2837614.2837625

[4] M. SHAPIRO, N. PREGUIÇA, C. BAQUERO, M. ZAWIRSKI.*Conflict-free Replicated Data Types*, in "Int. Symp. on Stabilization, Safety, and Security of Dist. Sys. (SSS)", Grenoble, France, X. DÉFAGO, F. PETIT, V. VILLAIN (editors), Lecture Notes in Comp. Sc., Springer-Verlag, October 2011, vol. 6976, p. 386–400, http://lip6.fr/Marc.Shapiro/papers/CRDTs_SSS-2011.pdf

[5] M. ZAWIRSKI, N. PREGUIÇA, S. DUARTE, A. BIENIUSA, V. BALEGAS, M. SHAPIRO.*Write Fast, Read in the Past: Causal Consistency for Client-side Applications*, in "Int. Conf. on Middleware (MIDDLEWARE)", Vancouver, BC, Canada, ACM/IFIP/Usenix, December 2015, p. 75–87

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[6] F. CORIAT.*Geolocation and communication in post-disaster situations*, Sorbonne Université, Faculté des Sciences et Ingénierie, December 2018, https://tel.archives-ouvertes.fr/tel-01970777

[7] L. HAMIDOUCHE.*Towards efficient dissemination of voluminous data over dense Wi-Fi networks*, Sorbonne Université, Faculté des Sciences et Ingénierie, June 2018, https://hal.inria.fr/tel-01953300

[8] D. JEANNEAU.*Failure Detectors in Dynamic Distributed Systems*, EDITE ; Sorbonne Université, December 2018, https://hal.archives-ouvertes.fr/tel-01951975

[9] B. NGOM.*FreeCore: a system for indexing document summaries on a Distributed Hash Table (DHT)*, Pierre and Marie Curie University, July 2018, https://hal.inria.fr/tel-01921587

[10] A. TOMSIC.*Exploring the design space of highly-available distributed transactions*, Sorbonne Universites, UPMC University of Paris 6, April 2018, https://hal.archives-ouvertes.fr/tel-01956321

[11] G. VORON.*Efficient Virtualization of NUMA Architectures*, Sorbonne Université, Faculté des Sciences et Ingénierie, March 2018, https://hal.archives-ouvertes.fr/tel-01947560

### Articles in International Peer-Reviewed Journal

[12] Z. AL-SHARA, F. ALVARES, H. BRUNELIERE, J. LEJEUNE, C. PRUD'HOMME, T. LEDOUX.*CoMe4ACloud: An End-to-End Framework for Autonomic Cloud Systems*, in "Future Generation Computer Systems", September 2018, vol. 86, p. 339-354 [*DOI :* 10.1016/J.FUTURE.2018.03.039], https://hal.archives-ouvertes.fr/hal-01762716

[13] L. BLIN, F. BOUBEKEUR, S. DUBOIS.*A Self-Stabilizing Memory Efficient Algorithm for the Minimum Diameter Spanning Tree under an Omnipotent Daemon*, in "Journal of Parallel and Distributed Computing", July 2018, vol. 117, p. 50-62 [*DOI :* 10.1016/J.JPDC.2018.02.007], https://hal.inria.fr/hal-01966265

[14] S. BOUCHARD, Y. DIEUDONNÉ, F. PETIT, A. PELC.*On Deterministic Rendezvous at a Node of Agents with Arbitrary Velocities*, in "Information Processing Letters", January 2018, vol. 133, p. 39 - 43 [*DOI :* 10.1016/J.IPL.2018.01.003], https://hal.archives-ouvertes.fr/hal-01701786

[15] A. GRACIELA DE MORAES ROSSETTO, C. GEYER, L. ARANTES, P. SENS.*Impact FD: An Unreliable Failure Detector Based on Process Relevance and Confidence in the System*, in "The Computer Journal", 2018, https://hal.inria.fr/hal-01793311

[16] L. A. RODRIGUES, E. P. DUARTE JR., L. ARANTES.*A distributed k-mutual exclusion algorithm based on autonomic spanning trees*, in "Journal of Parallel and Distributed Computing", May 2018, vol. 115, p. 41-55 [*DOI :* 10.1016/J.JPDC.2018.01.008], https://hal.inria.fr/hal-01965673

[17] J. P. DE ARAUJO, L. ARANTES, E. P. DUARTE JR., L. A. RODRIGUES, P. SENS.*VCube-PS: A causal broadcast topic-based publish/subscribe system*, in "Journal of Parallel and Distributed Computing", November 2018 [*DOI :* 10.1016/J.JPDC.2018.10.011], https://hal.inria.fr/hal-01925856

### Articles in Non Peer-Reviewed Journal

[18] A. BIENIUSA, A. GOTSMAN, B. KEMME, M. SHAPIRO.*Data Consistency in Distributed Systems: Algorithms, Programs, and Databases*, in "Dagstuhl Reports", July 2018, vol. 8, n$^{o}$ 2, p. 101-121 [*DOI :* 10.4230/DAGREP.8.2.101], https://hal.inria.fr/hal-01848384

### International Conferences with Proceedings

[19] L. ARANTES, E. BAMPIS, A. KONONOV, M. LETSIOS, G. LUCARELLI, P. SENS.*Scheduling under Uncertainty: A Query-based Approach*, in "IJCAI 2018 - 27th International Joint Conference on Artificial Intelligence", Stockholm, Sweden, July 2018, https://hal.inria.fr/hal-01924648

[20] S. BOUCHARD, M. BOURNAT, Y. DIEUDONNÉ, S. DUBOIS, F. PETIT.*Approche asynchrone dans le plan : un algorithme déterministe polynomial*, in "ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01782388

[21] S. BOUCHARD, Y. DIEUDONNÉ, B. DUCOURTHIAL.*Rassemblement byzantin dans les réseaux*, in "20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications(ALGOTEL 2018)", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01782387

[22] S. BOUCHARD, Y. DIEUDONNÉ, A. LAMANI.*Byzantine Gathering in Polynomial Time*, in "45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)", Prague, Czech Republic, July 2018 [*DOI :* 10.4230/LIPICS.ICALP.2018.147], https://hal.archives-ouvertes.fr/hal-01965743

[23] S. BOUCHARD, Y. DIEUDONNÉ, A. PELC, F. PETIT.*Deterministic Treasure Hunt in the Plane with Angular Hints*, in "29th International Symposium on Algorithms and Computation, ISAAC 2018", Jiaoxi Township, Taiwan, W.-L. HSU, D.-T. LEE, C.-S. LIAO (editors), Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, December 2018, vol. 123, p. 48:1–48:13 [*DOI :* 10.4230/LIPICS.ISAAC.2018.48], https://hal.sorbonne-universite.fr/hal-01970990

[24] J. BOURON, S. CHEVALLEY, B. LEPERS, W. ZWAENEPOEL, R. GOUICEM, J. LAWALL, G. MULLER, J. SOPENA.*The Battle of the Schedulers: FreeBSD ULE vs. Linux CFS*, in "2018 USENIX Annual Technical Conference", Boston, MA, United States, July 2018, https://hal.inria.fr/hal-01853267

[25] H. BRUNELIERE, Z. AL-SHARA, F. ALVARES, J. LEJEUNE, T. LEDOUX.*A Model-based Architecture for Autonomic and Heterogeneous Cloud Systems*, in "CLOSER 2018 - 8h International Conference on Cloud Computing and Services Science", Funchal, Portugal, March 2018, vol. 1, p. 201-212, Best Paper Award [*DOI :* 10.5220/0006773002010212], https://hal.archives-ouvertes.fr/hal-01705248

[26] G. FRAYSSE, I. GRIDA BEN YAHIA, J. LEJEUNE, P. SENS, J. SOPENA.*Towards multi-SDN services: Dangers of concurrent resource allocation from multiple providers*, in "21st Conference on Innovation in Clouds, Internet and Networks (ICIN 2018)", Paris, France, February 2018, https://hal.inria.fr/hal-01793636

[27] E. MAUFFRET, D. JEANNEAU, L. ARANTES, P. SENS.*The Weakest Failure Detector to Solve the Mutual Exclusion Problem in an Unknown Dynamic Environment*, in "20th International Conference on Distributed Computing and Networking (ICDCN 2019)", Bangalore, India, January 2019, Extended version: https://hal.archives-ouvertes.fr/hal-01661127v3, https://hal.archives-ouvertes.fr/hal-01929224

[28] M. NAJAFZADEH, M. SHAPIRO, P. EUGSTER.*Co-Design and Verification of an Available File System*, in "VMCAI 2018 - International Conference on Verification, Model Checking, and Abstract Interpretation", Los Angeles, CA, United States, I. DILLIG, J. PALSBERG (editors), Lecture Notes in Computer Science, Springer, January 2018, vol. 10747, p. 358-381 [*DOI :* 10.1007/978-3-319-73721-8_17], https://hal.inria.fr/hal-01696263

[29] L. A. RODRIGUES, E. P. DUARTE JR., J. P. DE ARAUJO, L. ARANTES, P. SENS.*Bundling Messages to Reduce the Cost of Tree-Based Broadcast Algorithms*, in "LADC 2018 - 8th Latin-American Symposium on Dependable Computing", Foz do Iguaçu, Brazil, October 2018, https://hal.inria.fr/hal-01959395

[30] M. SHAPIRO, A. BIENIUSA, P. ZELLER, G. PETRI.*Ensuring referential integrity under causal consistency*, in "PaPoC 2018 - 5th Workshop on Principles and Practice of Consistency for Distributed Data", Porto, Portugal, April 2018, https://arxiv.org/abs/1803.03482 , https://hal.inria.fr/hal-01727207

[31] A. Z. TOMSIC, M. BRAVO, M. SHAPIRO.*Distributed transactional reads: the strong, the quick, the fresh & the impossible*, in "2018 ACM/IFIP/USENIX International Middleware Conference", Rennes, France, Proceedings of 2018 ACM/IFIP/USENIX International Middleware Conference, ACM, December 2018, 14, https://arxiv.org/abs/1810.01698 [*DOI :* 10.1145/3274808.3274818], https://hal.inria.fr/hal-01876456

[32] D. VASILAS, M. SHAPIRO, B. KING.*A Modular Design for Geo-Distributed Querying: Work in Progress Report*, in "PaPoC 2018 - 5th Workshop on Principles and Practice of Consistency for Distributed Data", Porto, Portugal, April 2018, p. 1-8, https://arxiv.org/abs/1803.04141 , https://hal.inria.fr/hal-01728712

[33] J. P. DE ARAUJO, L. ARANTES, E. P. DUARTE JR., L. A. RODRIGUES, P. SENS.*A Communication-Efficient Causal Broadcast Protocol*, in "ICPP 2018 - 47th International Conference on Parallel Processing", Eugene, Oregon, United States, August 2018, https://hal.inria.fr/hal-01924741

### Conferences without Proceedings

[34] K. ALTISEN, S. DEVISMES, A. DURAND, F. PETIT.*Stabilisation progressive*, in "ALGOTEL 2018 - 20èmes Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications", Roscoff, France, May 2018, https://hal.archives-ouvertes.fr/hal-01779963

[35] F. CORIAT, A. FLADENMULLER, L. ARANTES.*Architecture de collecte pour la géolocalisation en situation de crise : évaluation comparative*, in "ComPAS 2018 - Conférence d'informatique en Parallélisme, Architecture et Système", Toulouse, France, July 2018, https://hal.sorbonne-universite.fr/hal-01877942

### Research Reports

[36] M. BOURNAT, S. DUBOIS, F. PETIT.*Gracefully Degrading Gathering in Dynamic Rings*, LIP6, Sorbonne Université, CNRS, UMR 7606 ; DELYS ; Inria, May 2018, https://arxiv.org/abs/1805.05137 , https://hal.archives-ouvertes.fr/hal-01790554

[37] E. MAUFFRET, D. JEANNEAU, L. ARANTES, P. SENS.*The Weakest Failure Detector to Solve the Mutual Exclusion Problem in an Unknown Dynamic Environment*, LISTIC ; Sorbonne Universités, UPMC Univ Paris 06, CNRS, LIP6 UMR 7606, October 2018 [*DOI :* 10.4230/LIPICS], https://hal.archives-ouvertes.fr/hal-01661127

[38] S. S. NAIR, M. SHAPIRO.*Improving the "Correct Eventual Consistency" Tool*, Sorbonne Université, July 2018, n$^{o}$ RR-9191, https://arxiv.org/abs/1807.06431 , https://hal.inria.fr/hal-01832888

[39] M. SHAPIRO, A. BIENIUSA, N. PREGUIÇA, V. BALEGAS, C. MEIKLEJOHN.*Just-Right Consistency: reconciling availability and safety*, Inria Paris ; UPMC - Paris 6 Sorbonne Universités ; Tech. U. Kaiserslautern ; U. Nova de Lisboa ; U. Catholique de Louvain, January 2018, n$^{o}$ RR-9145, p. 1-15, https://arxiv.org/abs/1801.06340 , https://hal.inria.fr/hal-01685945

### Scientific Popularization

[40] N. PREGUIÇA, C. BAQUERO, M. SHAPIRO.*Conflict-free Replicated Data Types (CRDTs)*, in "Encyclopedia of Big Data Technologies", Encyclopedia of Big Data Technologies, Springer International Publishing, April 2018, https://arxiv.org/abs/1805.06358 [*DOI :* 10.1007/978-3-319-63962-8_185-1], https://hal.inria.fr/hal-01793586

[41] M. SHAPIRO, P. SUTRA.*Database Consistency Models*, in "Encyclopedia of Big Data Technologies", S. SAK, A. ZOMAYA (editors), Springer, February 2018, https://arxiv.org/abs/1804.00914 [*DOI :* 10.1007/978-3-319-63962-8_203-1], https://hal.inria.fr/hal-01756780

### Other Publications

[42] B. NGOM, M. MAKPANGOU, S. NDIAYE.*SPT – Summary Prefix Tree: An over DHT Indexing Data Structure for Efficient Superset Search*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01757074

# Project-Team DYOGENE

## Dynamics of Geometric Networks

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Networks and Telecommunications**

# Table of contents

# Project-Team DYOGENE

*Creation of the Project-Team: 2013 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.2.4. - QoS, performance evaluation
A6.1.4. - Multiscale modeling
A6.2.3. - Probabilistic methods
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A8.3. - Geometry, Topology
A8.6. - Information theory
A8.7. - Graph theory
A8.8. - Network science
A8.9. - Performance evaluation
A9.2. - Machine learning
A9.7. - AI algorithmics

**Other Research Topics and Application Domains:**

B4.3. - Renewable energy production
B6.2.2. - Radio technology
B6.3.4. - Social Networks

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Bartlomiej Blaszczyszyn [Inria, Senior Researcher, HDR]
Marc Lelarge [Team leader, Inria, Senior Researcher, HDR]
François Baccelli [Inria, Senior Researcher, HDR]
Ana Busic [Inria, Researcher]
Christine Fricker [Inria, Researcher, from Mar 2018]
Laurent Massoulié [Inria, Senior Researcher, from Feb 2018, HDR]

**External Collaborators**

Anne Bouillard [Nokia, HDR]
Pierre Bremaud [EPFL]

**Technical Staff**

Holger Keeler [Inria, granted by HUAWEI TECHNOLOGIES FRANCE]

**PhD Students**

Arnaud Cadas [PSL]
Alexis Galland [Ministère des Armées]
Md Umar Hashmi [PSL]
Hadrien Hendrikx [Inria, from Oct 2018]
Dalia-Georgiana Popescu [Bell Labs (Alcatel), until Oct 2018]
Alexandre Hollocou [Ministère de la Défense, until Nov 2018]
Quentin Le Gall [Inria]

Leo Miolane [Ecole polytechnique]
Edouard Pineau [Safran]
Sébastien Samain [Inria]
Ludovic Stephan [Ecole Normale Supérieure Paris, from Sep 2018]
Rémi Varloot [Inria, until Feb 2018]

**Post-Doctoral Fellow**

Cheng Wan [Université Paris-Sud and Inria, granted by PGMO, from Feb 2018 until Nov 2018]

**Visiting Scientists**

Akshay Goel [Feb 2018]
Sean Meyn [University of Florida, from Jul 2018 until Dec 2018]

**Administrative Assistants**

Helene Bessin Rousseau [Inria, from Mar 2018]
Helene Milome [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

The general scientific focus of DYOGENE is on the development of network mathematics. The following theories lie within our research interest: dynamical systems, queuing theory, optimization and control, information theory, stochastic processes, random graphs, stochastic geometry.

Our theoretical developments are motivated by and applied in the context of communication networks (Internet, wireless, mobile, cellular, peer-to-peer), social and economic networks, power grids.

We collaborate with many industrial partners. Our current industrial relations involve EDF, Google, Huawei, Microsoft, Nokia, Orange, Safran.

More specifically, the scientific focus of DYOGENE defined in 2013 was on geometric network dynamics arising in communications. By geometric networks we understand networks with a nontrivial, discrete or continuous, geometric definition of the existence of links between the nodes. In stochastic geometric networks, this definition leads to random graphs or stochastic geometric models.

A first type of geometric network dynamics is the one where the nodes or the links change over time according to an exogeneous dynamics (e.g. node motion and geometric definition of the links). We will refer to this as dynamics of geometric networks below. A second type is that where links and/or nodes are fixed but harbor local dynamical systems (in our case, stemming from e.g. information theory, queuing theory, social and economic sciences). This will be called dynamics on geometric networks. A third type is that where the dynamics of the network geometry and the local dynamics interplay. Our motivations for studying these systems stem from many fields of communications where they play a central role, and in particular: message passing algorithms; epidemic algorithms; wireless networks and information theory; device to device networking; distributed content delivery; social and economic networks, power grids.

# 3. Research Program

## 3.1. Initial research axes

The following research axes have been defined in 2013 when the project-team was created.

- Algorithms for network performance analysis, led by A. Bouillard and A. Busic.
- Stochastic geometry and information theory for wireless network, led by B. Blaszczyszyn and F. Baccelli.
- The cavity method for network algorithms, led by M. Lelarge.

Our scientific interests keep evolving. Research areas which received the most of our attention in 2017 are summarized in the following sections.

## 3.2. Distributed network control and smart-grids

Foundation of an entirely new science for distributed control of networks with applications to the stabilization of power grids subject to high volatility of renewable energy production is being developed A. Busic in collaboration with A. Bouillard and Sean Meyn [University of Florida].

## 3.3. Mathematics of wireless cellular networks

A comprehensive approach involving information theory, queueing and stochastic geometry to model and analyze the performance of large cellular networks, validated and implemented by Orange is being led by B. Blaszczyszyn in collaboration with F. Baccelli and M. K. Karray [Orange Labs]

## 3.4. High-dimensional statistical inference for social networks

Community detection and non-regular ramanujan graphs sole a conjecture on the optimality of non-backtracking spectral algorithm for community dectection in sparse stochastic block model graphs, as has been proved by M. Lelarge and L. Massoulié in collaboration with C. Bordenave [IMT Toulouse].

# 4. Application Domains

## 4.1. Physical communication networks

Internet, wireless, mobile, cellular networks.

## 4.2. Abstract networks

Social interactions, human communities, economic networks.

## 4.3. Power grids

Energy networks.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

Publication of a monograph *Stochastic Geometry Analysis of Cellular Networks* by Cambridge University Press [30] that presents latest analytic techniques and results from stochastic geometry for modelling of heterogeneous cellular networks.

### *5.1.1. Awards*

Our paper “ Optimal Algorithms for Non-Smooth Distributed Optimization in Networks” by K. Scaman, F. Bach, S. Bubeck, Y.T. Lee and L. Massoulié won a best paper award at the NeurIPS 2018 conference.

# 6. New Results

## 6.1. Energy Trade-offs for end-to-end Communications in Urban Vehicular Networks exploiting an Hyperfractal Model

In [33] presented this year at MSWIM DIVANet we show results on the trade-offs between the end-to-end communication delay and energy spent for completing a transmission in vehicular communications in urban settings. This study exploits our innovative model called " hyperfractal " that captures the self-similarity of the topology and vehicle locations in cities. We enrich the model by incorporating roadside infrastructure. We use analytical tools to derive theoretical bounds for the end-to-end communication hop count under two different energy constraints: either total accumulated energy, or maximum energy per node. More precisely, we prove that the hop count is bounded by $O(n(1-\alpha)/(dm-1))$ where $\alpha < 1$ and $m > 2$ is the precise hyperfractal dimension. This proves that for both constraints the energy decreases as we allow to chose among paths of larger length. In fact the asymptotic limit of the energy becomes significantly small when the number of nodes becomes asymptotically large. A lower bound on the network throughput capacity with constraints on path energy is also given. The results are confirmed through exhaustive simulations using different hyperfractal dimensions and path loss coefficients.

## 6.2. Broadcast Speedup in Vehicular Networks via Information Teleportation

In [32] presented this year at LCN our goal is to increase our understanding of the fundamental communication properties in urban vehicle-to-vehicle mobile networks by exploiting the self-similarity and hierarchical organization of modern cities. We use an innovative model called "hyperfractal" that captures the self-similarities of both the traffic and vehicle locations, and yet avoids the extremes of regularity and randomness. We use analytical tools to derive matching theoretical upper and lower bounds for the information propagation speed in an urban delay tolerant network (i.e., a network that is disconnected at all time, and thus uses a store-carry-and-forward routing model). We prove that the average broadcast time behaves as $n(1-\delta)$ (times a slowly varying function), where $\delta$ depends on the precise fractal dimension. Furthermore, we show that the broadcast speedup is due in part to an interesting self-similar phenomenon, that we denote as information teleportation. This phenomenon arises as a consequence of the topology of the vehicle traffic, and triggers an acceleration of the broadcast time. We show that our model fits real cities where open traffic data sets are available. The study presents simulations that confirm the validity of the bounds in multiple realistic settings, including scenarios with variable speed.

## 6.3. Vehicle-to-Infrastructure Communications Design in Urban Hyperfractals

In [25] presented at SPAWC our goal is to increase the awareness about the communication opportunities that arise in urban vehicle networks when exploiting the self-similarity and hierarchical organization of modern cities. The work uses our innovative model called " hyperfractal " that captures the self-similarity of the urban vehicular networks as well as incorporating roadside infrastructure with its own self-similarity. We use analytical tools to provide achievable trade-offs in operating the roadside units under the constraint of minimum routing path delay while maintaining a reasonably balanced load. The models and results are supported by simulations with different city hyperfractal dimensions in two different routing scenarios: nearest neighbor routing with no collision and minimum delay routing model assuming slotted Aloha, signal to interference ratio (SIR) capture condition, power-path loss, Rayleigh fading.

## 6.4. Book on Stochastic Geometry Analysis of Cellular Networks

In 2018 we have published a monograph [30] in which we explain the very latest analytic techniques and results from stochastic geometry for modelling the signal-to-interference-plus-noise ratio (SINR) distribution in heterogeneous cellular networks. This book is supposed to help readers to understand the effects of combining different system deployment parameters on key performance indicators such as coverage and

capacity, enabling the efficient allocation of simulation resources. In addition to covering results for network models based on the Poisson point process, this book presents recent results for when non-Poisson base station configurations appear Poisson, due to random propagation effects such as fading and shadowing, as well as non-Poisson models for base station configurations, with a focus on determinantal point processes and tractable approximation methods. Theoretical results are illustrated with practical Long-Term Evolution (LTE) applications and compared with real-world deployment results.

## 6.5. Gibbsian On-Line Distributed Content Caching Strategy for Cellular Networks

In [9], we develop Gibbs sampling based techniques for learning the optimal content placement in a cellular network. A collection of base stations are scattered on the space, each having a cell (possibly overlapping with other cells). Mobile users request for downloads from a finite set of contents according to some popularity distribution. Each base station can store only a strict subset of the contents at a time; if a requested content is not available at any serving base station, it has to be downloaded from the backhaul. Thus, there arises the problem of optimal content placement which can minimize the download rate from the backhaul, or equivalently maximize the cache hit rate. Using similar ideas as Gibbs sampling, we propose simple sequential content update rules that decide whether to store a content at a base station based on the knowledge of contents in neighbouring base stations. The update rule is shown to be asymptotically converging to the optimal content placement for all nodes. Next, we extend the algorithm to address the situation where content popularities and cell topology are initially unknown, but are estimated as new requests arrive to the base stations. Finally, improvement in cache hit rate is demonstrated numerically.

## 6.6. Location Aware Opportunistic Bandwidth Sharing between Static and Mobile Users with Stochastic Learning in Cellular Networks

In [7], we consider location-dependent opportunistic bandwidth sharing between static and mobile downlink users in a cellular network. Each cell has some fixed number of static users. Mobile users enter the cell, move inside the cell for some time and then leave the cell. In order to provide higher data rate to mobile users, we propose to provide higher bandwidth to the mobile users at favourable times and locations, and provide higher bandwidth to the static users in other times. We formulate the problem as a long run average reward Markov decision process (MDP) where the per-step reward is a linear combination of instantaneous data volumes received by static and mobile users, and find the optimal policy. The transition structure of this MDP is not known in general. To alleviate this issue, we propose a learning algorithm based on single timescale stochastic approximation. Also, noting that the unconstrained MDP can be used to solve a constrained problem, we provide a learning algorithm based on multi-timescale stochastic approximation. The results are extended to address the issue of fair bandwidth sharing between the two classes of users. Numerical results demonstrate performance improvement by our scheme, and also the trade-off between performance gain and fairness.

## 6.7. Performance analysis of cellular networks with opportunistic scheduling using queueing theory and stochastic geometry

In [38] submitted this year, combining stochastic geometric approach with some classical results from queuing theory,we propose a comprehensive framework for the performance study of large cellular networks featuring opportunistic scheduling. Rapid and verifiable with respect to real data, our approach is particularly useful for network dimensioning and long term economic planning. It is based on a detailed network model combining an information-theoretic representation of the link layer, a queuing-theoretic representation of the users' scheduler, and a stochastic-geometric representation of the signal propagation and the network cells. It allows one to evaluate principal characteristics of the individual cells, such as loads (defined as the fraction of time the cell is not empty), the mean number of served users in the steady state, and the user throughput. A simplified Gaussian approximate model is also proposed to facilitate study of the spatial distribution of these metrics across the network. The analysis of both models requires only simulations of the point process of base stations

and the shadowing field to estimate the expectations of some stochastic-geometric functionals not admitting explicit expressions. A key observation of our approach , bridging spatial and temporal analysis, relates the SINR distribution of the typical user to the load of the typical cell of the network. The former is a static characteristic of the network related to its spectral efficiency while the latter characterizes the performance of the (generalized) processor sharing queue serving the dynamic population of users of this cell.

## 6.8. The Influence of Canyon Shadowing on Device-to-Device Connectivity in Urban Scenario

In [48] submitted this year, we use percolation theory to study the feasibility of large-scale connectivity of relay-augmented device-to-device (D2D) networks in an urban scenario, featuring a haphazard system of streets and canyon shadowing allowing only for line-of-sight (LOS) communications in a limited finite range. We use a homogeneous Poisson-Voronoi tessellation (PVT) model of streets with homogeneous Poisson users (devices) on its edges and independent Bernoulli relays on the vertices. Using this model, we demonstrated the existence of a minimal threshold for relays below which large-scale connectivity of the network is not possible, regardless of all other network parameters. Through simulations, we estimated this threshold to 71.3%. Moreover, if the mean street length is not larger than some threshold (predicted to 74.3% of the communication range; which might be the case in a typical urban scenario) then any (whatever small) density of users can be compensated by equipping more crossroads with relays. Above this latter threshold, good connectivity requires some minimal density of users, compensated by the relays in a way we make explicit. The existence of the above regimes brings interesting qualitative arguments to the discussion on the possible D2D deployment scenarios.

## 6.9. Determinantal thinning of point processes with network learning applications

In [39] submitted this year, a new type of dependent thinning for point processes in continuous space is proposed, which leverages the advantages of determinantal point processes defined on finite spaces and, as such, is particularly amenable to statistical, numerical, and simulation techniques. It gives a new point process that can serve as a network model exhibiting repulsion. The properties and functions of the new point process, such as moment measures, the Laplace functional, the void probabilities, as well as conditional (Palm) characteristics can be estimated accurately by simulating the underlying (non-thinned) point process, which can be taken, for example, to be Poisson. This is in contrast (and preference to) finite Gibbs point processes, which, instead of thinning, require weighting the Poisson realizations, involving usually intractable normalizing constants. Models based on determinantal point processes are also well suited for statistical (supervised) learning techniques, allowing the models to be fitted to observed network patterns with some particular geometric properties. We illustrate this approach by imitating with determinantal thinning the well-known Matérn II hard-core thinning, as well as a soft-core thinning depending on nearest-neighbour triangles. These two examples demonstrate how the proposed approach can lead to new, statistically optimized, probabilistic transmission scheduling schemes.

## 6.10. Analyzing LoRa long-range, low-power, wide-area networks using stochastic geometry

In [40] submitted this year, we present a simple, stochastic-geometric model of a wireless access network exploiting the LoRA (Long Range) protocol, which is a non-expensive technology allowing for long-range, single-hop connectivity for the Internet of Things. We assume a space-time Poisson model of packets transmitted by LoRA nodes to a fixed base station. Following previous studies of the impact of interference, we assume that a given packet is successfully received when no interfering packet arrives with similar power before the given packet payload phase. This is as a consequence of LoRa using different transmission rates for different link budgets (transmissions with smaller received powers use larger spreading factors) and LoRa intra-technology interference treatment. Using our model, we study the scaling of the packet reception

probabilities per link budget as a function of the spatial density of nodes and their rate of transmissions. We consider both the parameter values recommended by the LoRa provider, as well as proposing LoRa tuning to improve the equality of performance for all link budgets. We also consider spatially non-homogeneous distributions of LoRa nodes. We show also how a fair comparison to non-slotted Aloha can be made within the same framework.

## 6.11. Statistical learning of geometric characteristics of wireless networks

In [41] to appear in Proc. INFOCOM 2019, motivated by the prediction of cell loads in cellular networks, we formulate the following new, fundamental problem of statistical learning of geometric marks of point processes: An unknown marking function, depending on the geometry of point patterns, produces characteristics (marks) of the points. One aims at learning this function from the examples of marked point patterns in order to predict the marks of new point patterns. To approximate (interpolate) the marking function, in our baseline approach, we build a statistical regression model of the marks with respect some local point distance representation. In a more advanced approach, we use a global data representation via the scattering moments of random measures, which build informative and stable to deformations data representation, already proven useful in image analysis and related application domains. In this case, the regression of the scattering moments of the marked point patterns with respect to the non-marked ones is combined with the numerical solution of the inverse problem, where the marks are recovered from the estimated scattering moments. Considering some simple, generic marks, often appearing in the modeling of wireless networks, such as the shot-noise values, nearest neighbour distance, and some characteristics of the Voronoi cells, we show that the scattering moments can capture similar geometry information as the baseline approach, and can reach even better performance, especially for non-local marking functions. Our results motivate further development of statistical learning tools for stochastic geometry and analysis of wireless networks, in particular to predict cell loads in cellular networks from the locations of base stations and traffic demand.

## 6.12. Ressource allocation in bike sharing systems

Vehicle sharing systems are becoming an urban mode of transportation, and launched in many cities, as Velib' and Autolib' in Paris. Managing such systems is quite difficult. One of the major issues is the availability of the resources: vehicles or free slots. These systems became a hot topic in Operation Research and the importance of stochasticity on the system behavior leads us to propose mathematical stochastic models. The aim is to understand the system behavior and how to manage these systems in order to improve the allocation of both resources to users.

To improve BSS (bike-sharing systems), two types of policies can be deployed: incentives to the users to choose a better station, called *natural* or *green* regulation, or redistribution by trucks, called *active* regulation. In a simple mathematical model, we proved the efficiency of the 2-choice incentive policy for BSS (bike-sharing systems). The drawback of the model is that it ignores the geometry of the system, where the choice is only local. The purpose of this first work is to deal with this policy in real systems.

We use data trip data obtained from JCDecaux and reports on station status collected as open data, to test local choice policy. Indeed we designed and tested a new policy relying on a local small change in user behaviors, by adapting their trips to resource availability around their departure and arrival stations, based on 2-choice policy. Results show that, even with a small user collaboration, the proposed method increases significantly the global balance of the bike sharing system and therefore the user satisfaction. This is done using trip data sets and detecting spatial outliers, stations having a behavior significantly different from their spatial neighbors, in a context where neighbors are heavily correlated. For that we proposed an improved version of the well-known Moran scatterplot method, using a robust distance metric called Gower similarity. Using this new version of Moran scatterplot, we show that, for the occupancy data set obtained by modifiying trips, the number of spatial outliers drastically decreases. We generalize this study with W. Ghanem and L. Massoulié testing incentive and redistribution policies on a simulator, where the tradeoff between the number of frustrated trips and the penalty for the users can be measured. We propose new versions of these policies including prediction.

## 6.13. Analyzing the choice of the least loaded queue between two neighboring queues

A model of N queues, with a local choice policy, is studied. Each one-server queue has a Poissonian arrival of customers. When a customer arrives at a queue, he joins the least loaded queue between this queue and the next one, ties solved at random. Service times have exponential distribution. The system is stable if the arrival-to-service rate ratio, also called load, is less than one. When the load tends to zero, we derive the first terms of the expansion in this parameter for the stationary probabilities that a queue has few customers. Then we provide explicit asymptotics, as the load tends to zero, for the stationary probabilities of the queue length. We used the analyticity of the stationary probabilities as a function of the load. It shows the behavior difference between this local choice policy and the 2-choice policy (*supermarket model*).

## 6.14. Optimal Content Replication and Request Matching in Large Caching Systems

We consider models of content delivery networks in which the servers are constrained by two main resources: memory and bandwidth. In such systems, the throughput crucially depends on how contents are replicated across servers and how the requests of specific contents are matched to servers storing those contents. In this paper, we first formulate the problem of computing the optimal replication policy which if combined with the optimal matching policy maximizes the throughput of the caching system in the stationary regime. It is shown that computing the optimal replication policy for a given system is an NP-hard problem. A greedy replication scheme is proposed and it is shown that the scheme provides a constant factor approximation guarantee. We then propose a simple randomized matching scheme which avoids the problem of interruption in service of the ongoing requests due to re-assignment or repacking of the existing requests in the optimal matching policy. The dynamics of the caching system is analyzed under the combination of proposed replication and matching schemes. We study a limiting regime, where the number of servers and the arrival rates of the contents are scaled proportionally, and show that the proposed policies achieve asymptotic optimality. Extensive simulation results are presented to evaluate the performance of different policies and study the behavior of the caching system under different service time distributions of the requests.

## 6.15. Statistical thresholds for Tensor PCA

This is a joint work with Aukosh Jagannath and Patrick Lopatto. We study the statistical limits of testing and estimation for a rank one deformation of a Gaussian random tensor. We compute the sharp thresholds for hypothesis testing and estimation by maximum likelihood and show that they are the same. Furthermore, we find that the maximum likelihood estimator achieves the maximal correlation with the planted vector among measurable estimators above the estimation threshold. In this setting, the maximum likelihood estimator exhibits a discontinuous BBP-type transition: below the critical threshold the estimator is orthogonal to the planted vector, but above the critical threshold, it achieves positive correlation which is uniformly bounded away from zero.

## 6.16. The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning

This is a joint work with Andrea Montanari. The Lasso is a popular regression method for high-dimensional problems in which the number of parameters $\theta_1, \cdots, \theta_N$, is larger than the number $n$ of samples: $N > n$. A useful heuristics relates the statistical properties of the Lasso estimator to that of a simple soft-thresholding denoiser, in a denoising problem in which the parameters $(\theta_i)_{i \leq N}$ are observed in Gaussian noise, with a carefully tuned variance. Earlier work confirmed this picture in the limit $n, N \to \infty$, pointwise in the parameters $\theta$, and in the value of the regularization parameter.

Here, we consider a standard random design model and prove exponential concentration of its empirical distribution around the prediction provided by the Gaussian denoising model. Crucially, our results are uniform with respect to $\theta$ belonging to $\ell_q$ balls, $q \in [0,1]$, and with respect to the regularization parameter. This allows to derive sharp results for the performances of various data-driven procedures to tune the regularization.

Our proofs make use of Gaussian comparison inequalities, and in particular of a version of Gordon's minimax theorem developed by Thrampoulidis, Oymak, and Hassibi, which controls the optimum value of the Lasso optimization problem. Crucially, we prove a stability property of the minimizer in Wasserstein distance, that allows to characterize properties of the minimizer itself.

## 6.17. Phase transitions in spiked matrix estimation: information-theoretic analysis

We study here the so-called spiked Wigner and Wishart models, where one observes a low-rank matrix perturbed by some Gaussian noise. These models encompass many classical statistical tasks such as sparse PCA, submatrix localization, community detection or Gaussian mixture clustering. The goal of these notes is to present in a unified manner recent results (as well as new developments) on the information-theoretic limits of these spiked matrix/tensor models. We compute the minimal mean squared error for the estimation of the low-rank signal and compare it to the performance of spectral estimators and message passing algorithms. Phase transition phenomena are observed: depending on the noise level it is either impossible, easy (i.e. using polynomial-time estimators) or hard (information-theoretically possible, but no efficient algorithm is known to succeed) to recover the signal.

## 6.18. Accelerated decentralized optimization with local updates for smooth and strongly convex objectives

We study the problem of minimizing a sum of smooth and strongly convex functions split over the nodes of a network in a decentralized fashion. We propose the algorithm $ESDACD$, a decentralized accelerated algorithm that only requires local synchrony. Its rate depends on the condition number $\kappa$ of the local functions as well as the network topology and delays. Under mild assumptions on the topology of the graph, $ESDACD$ takes a time $O((\tau_{\max} + \Delta_{\max})\sqrt{\kappa/\gamma}\ln(\epsilon^{-1}))$ to reach a precision $\epsilon$ where $\gamma$ is the spectral gap of the graph, $\tau_{\max}$ the maximum communication delay and $\Delta_{\max}$ the maximum computation time. Therefore, it matches the rate of $SSDA$, which is optimal when $\tau_{\max} = \Omega(\Delta_{\max})$. Applying $ESDACD$ to quadratic local functions leads to an accelerated randomized gossip algorithm of rate $O(\sqrt{\theta_{\text{gossip}}/n})$ where $\theta_{\text{gossip}}$ is the rate of the standard randomized gossip. To the best of our knowledge, it is the first asynchronous gossip algorithm with a provably improved rate of convergence of the second moment of the error. We illustrate these results with experiments in idealized settings.

## 6.19. Group synchronization on grids

Group synchronization requires to estimate unknown elements $(\theta_v)_{v\in V}$ of a compact group $\mathbb{G}$ associated to the vertices of a graph $G = (V, E)$, using noisy observations of the group differences associated to the edges. This model is relevant to a variety of applications ranging from structure from motion in computer vision to graph localization and positioning, to certain families of community detection problems.

We focus on the case in which the graph $G$ is the $d$-dimensional grid. Since the unknowns $\theta_v$ are only determined up to a global action of the group, we consider the following weak recovery question. Can we determine the group difference $\theta_u^{-1}\theta_v$ between far apart vertices $u, v$ better than by random guessing? We prove that weak recovery is possible (provided the noise is small enough) for $d \geq 3$ and, for certain finite groups, for $d \geq 2$. Vice-versa, for some continuous groups, we prove that weak recovery is impossible for $d = 2$. Finally, for strong enough noise, weak recovery is always impossible.

## 6.20. An Impossibility Result for Reconstruction in a Degree-Corrected Planted-Partition Model

We consider the Degree-Corrected Stochastic Block Model (DC-SBM): a random graph on n nodes, having i.i.d. weights $(\phi_u)_{u=1}^n$ (possibly heavy-tailed), partitioned into $q \geq 2$ asymptotically equal-sized clusters. The model parameters are two constants $a, b > 0$ and the finite second moment of the weights $\Phi^{(2)}$. Vertices $u$ and $v$ are connected by an edge with probability $(\phi_u\phi_v/n)a$ when they are in the same class and with probability $(\phi_u\phi_v/n)b$ otherwise. We prove that it is information-theoretically impossible to estimate the clusters in a way positively correlated with the true community structure when $(a-b)2\Phi^{(2)} \leq q(a+b)$. As by-products of our proof we obtain (1) a precise coupling result for local neighbourhoods in DC-SBM's, that we use in a follow up paper [Gulikers et al., 2017] to establish a law of large numbers for local-functionals and (2) that long-range interactions are weak in (power-law) DC-SBM's.

## 6.21. On the capacity of information processing systems

We propose and analyze a family of information processing systems, where a finite set of experts or servers are employed to extract information about a stream of incoming jobs. Each job is associated with a hidden label drawn from some prior distribution. An inspection by an expert produces a noisy outcome that depends both on the job's hidden label and the type of the expert, and occupies the expert for a finite time duration. A decision maker's task is to dynamically assign inspections so that the resulting outcomes can be used to accurately recover the labels of all jobs, while keeping the system stable. Among our chief motivations are applications in crowd-sourcing, diagnostics, and experiment designs, where one wishes to efficiently learn the nature of a large number of items, using a finite pool of computational resources or human agents. We focus on the capacity of such an information processing system. Given a level of accuracy guarantee, we ask how many experts are needed in order to stabilize the system, and through what inspection architecture. Our main result provides an adaptive inspection policy that is asymptotically optimal in the following sense: the ratio between the required number of experts under our policy and the theoretical optimal converges to one, as the probability of error in label recovery tends to zero.

## 6.22. Optimal Algorithms for Non-Smooth Distributed Optimization in Networks

In this work, we consider the distributed optimization of non-smooth convex functions using a network of computing units. We investigate this problem under two regularity assumptions: (1) the Lipschitz continuity of the global objective function, and (2) the Lipschitz continuity of local individual functions. Under the local regularity assumption, we provide the first optimal first-order decentralized algorithm called multi-step primal-dual (MSPD) and its corresponding optimal convergence rate. A notable aspect of this result is that, for non-smooth functions, while the dominant term of the error is in $O(1/\sqrt{t})$, the structure of the communication network only impacts a second-order term in $O(1/t)$, where $t$ is time. In other words, the error due to limits in communication resources decreases at a fast rate even in the case of non-strongly-convex objective functions. Under the global regularity assumption, we provide a simple yet efficient algorithm called distributed randomized smoothing (DRS) based on a local smoothing of the objective function, and show that DRS is within a $d^{1/4}$ multiplicative factor of the optimal convergence rate, where $d$ is the underlying dimension.

## 6.23. Zap Meets Momentum: Stochastic Approximation Algorithms with Optimal Convergence Rate

There are two well known Stochastic Approximation techniques that are known to have optimal rate of convergence (measured in terms of asymptotic variance): the Ruppert-Polyak averaging technique, and stochastic Newton-Raphson (SNR)(a matrix gain algorithm that resembles the deterministic Newton-Raphson method). The Zap algorithms, introduced by Devraj and Meyn in 2017, are a version of SNR designed

to behave more closely like their deterministic cousin. It is found that estimates from the Zap Q-learning algorithm converge remarkably quickly, but the per-iteration complexity can be high. In [43], we introduce a new class of stochastic approximation algorithms based on matrix momentum. For a special choice of the matrix momentum and gain sequences, it is found in simulations that the parameter estimates obtained from the algorithm couple with those obtained from the more complex stochastic Newton-Raphson algorithm. Conditions under which coupling is guaranteed are established for a class of linear recursions. Optimal finite-$n$ error bounds are also obtained.

## 6.24. Ergodic theory for controlled Markov chains with stationary inputs

Consider a stochastic process $\mathbf{X}$ on a finite state space $X = \{1, \cdots, d\}$. It is conditionally Markov, given a real-valued 'input process' $\zeta$. This is assumed to be small, which is modeled through the scaling, $\zeta_t = \varepsilon \zeta^1_t$, $0 \le \varepsilon \le 1$, where $\zeta^1$ is a bounded stationary process. The following conclusions are obtained, subject to smoothness assumptions on the controlled transition matrix and a mixing condition on $\zeta$:

- A stationary version of the process is constructed, that is coupled with a stationary version of the Markov chain $\mathbf{X}^\bullet$ obtained with $\zeta \equiv 0$. The triple $(\mathbf{X}, \mathbf{X}^\bullet, \zeta)$ is a jointly stationary process satisfying $\mathsf{P}\{X(t) \neq X^\bullet(t)\} = O(\varepsilon)$. Moreover, a second-order Taylor-series approximation is obtained:

$$\mathsf{P}\{X(t) = i\} = \mathsf{P}\{X^\bullet(t) = i\} + \varepsilon^2 \varrho(i) + o(\varepsilon^2), \quad 1 \le i \le d,$$

  with an explicit formula for the vector $\varrho \in \mathfrak{R}^d$.

- For any $m \ge 1$ and any function $f : \{1, \cdots, d\} \times \mathfrak{R} \to \mathfrak{R}^m$, the stationary stochastic process $Y(t) = f(X(t), \zeta(t))$ has a power spectral density $\mathrm{S}_f$ that admits a second order Taylor series expansion: A function $\mathrm{S}^{(2)}{}_f : [-\pi, \pi] \to C^{m \times m}$ is constructed such that

$$\mathrm{S}_f(\theta) = \mathrm{S}^\bullet_f(\theta) + \varepsilon^2 \mathrm{S}^{(2)}{}_f(\theta) + o(\varepsilon^2), \quad \theta \in [-\pi, \pi].$$

  An explicit formula for the function $\mathrm{S}^{(2)}{}_f$ is obtained, based in part on the bounds in (i).

The results are illustrated using a version of the timing channel of Anantharam and Verdu.

## 6.25. Ordinary Differential Equation Methods for Markov Decision Processes and Application to Kullback–Leibler Control Cost

A new approach to computation of optimal policies for MDP (Markov decision process) models is introduced in [5], published in SICON this year. The main idea is to solve not one, but an entire family of MDPs, parameterized by a scalar $\zeta$ that appears in the one-step reward function. For an MDP with $d$ states, the family of relative value functions $\{h^*_\zeta : \zeta \in \mathbb{R}\}$ is the solution to an ODE, $\frac{d}{d\zeta} h^*_\zeta = \mathcal{V}(h^*_\zeta)$, where the vector field $\mathcal{V} : R^d \to R^d$ has a simple form, based on a matrix inverse. Two general applications are presented: Brockett's quadratic-cost MDP model, and a generalization of the "linearly solvable" MDP framework of Todorov in which the one-step reward function is defined by Kullback–Leibler divergence with respect to nominal dynamics. This technique was introduced by Todorov in 2007, where it was shown under general conditions that the solution to the average-reward optimality equations reduce to a simple eigenvector problem. Since then many authors have sought to apply this technique to control problems and models of bounded rationality in economics. A crucial assumption is that the input process is essentially unconstrained. For example, if the nominal dynamics include randomness from nature (eg, the impact of wind on a moving vehicle), then the optimal control solution does not respect the exogenous nature of this disturbance. In [16] we introduce a technique to solve a more general class of action-constrained MDPs.

## 6.26. Distributed control design for balancing the grid using flexible loads

Inexpensive energy from the wind and the sun comes with unwanted volatility, such as ramps with the setting sun or a gust of wind. Controllable generators manage supply-demand balance of power today, but this is becoming increasingly costly with increasing penetration of renewable energy. It has been argued since the 1980s that consumers should be put in the loop: "demand response" will help to create needed supply-demand balance. However, consumers use power for a reason and expect that the quality of service (QoS) they receive will lie within reasonable bounds. Moreover, the behavior of some consumers is unpredictable, while the grid operator requires predictable controllable resources to maintain reliability.

The goal of the book chapter [31] is to describe an emerging science for demand dispatch that will create virtual energy storage from flexible loads. By design, the grid-level services from flexible loads will be as controllable and predictable as a generator or fleet of batteries. Strict bounds on QoS will be maintained in all cases. The potential economic impact of these new resources is enormous. California plans to spend billions of dollars on batteries that will provide only a small fraction of the balancing services that can be obtained using demand dispatch. The potential impact on society is enormous: a sustainable energy future is possible with the right mix of infrastructure and control systems.

In [17], presented at IEEE CDC 2018, a natural notion of *energy capacity* is proposed for the special case of thermostatically controlled loads (TCLs). It is shown that this quantity is closely approximated by thermal energy capacity, which is a component of the "leaky battery model" introduced in prior work. Simulation experiments in a distributed control setting show that these energy limits, and accompanying power capacity limits, are reliable indicators of online capacity, even for a heterogeneous population of loads. A feedforward/feedback control scheme is proposed for a large collection of heterogeneous loads. At the local level, control loops are used to create cooperative responses from each load in a given class of homogeneous loads. This simplifies control of the aggregate based on two pieces of information: aggregate power consumption from each class of loads and the *state of charge* surrogate that is a part of the leaky battery model. This information is required at a slow time-scale (say, 5 minute sampling).

In [18], we study the problem of coordination of a collection of on/off thermostatically controlled loads (TCLs) to act as a "virtual battery". Virtual Energy Storage (VES) is provided by the collection by either consuming more (charging) or less (discharging) power than the baseline. VES can be an inexpensive alternative to batteries when a large share of the electricity comes from volatile sources such as solar and wind. Almost all prior work has assumed that the outside weather - which significantly effects a TCLs behavior - is constant. We combine the above distributed load control design with a grid level MPC (model predictive control) that uses predictions of disturbances (weather) over a planning horizon. Additionally, irrespective of the choice of control architecture, there is a fundamental limit to the power and energy capacity of the collection of TCLs. We partially address this issue by scaling the reference signal by a function of the outside air temperature.

## 6.27. Estimation and control of quality of service in demand dispatch

Flexibility of energy consumption can be harnessed for the purposes of grid-level ancillary services. In particular, through distributed control of a collection of loads, a balancing authority regulation signal can be tracked accurately, while ensuring that the quality of service (QoS) for each load is acceptable on average. Subject to distributed control approaches advocated in recent research, the histogram of QoS is approximately Gaussian, and consequently, each load will eventually receive poor service. In [11], published this year in IEEE Transactions on Smart Grid, statistical techniques are developed to estimate the mean and variance of QoS as a function of the power spectral density of the regulation signal. It is also shown that additional local control can eliminate risk. The histogram of QoS is truncated through this local control, so that strict bounds on service quality are guaranteed. While there is a tradeoff between the grid-level tracking performance (capacity and accuracy) and the bounds imposed on QoS, it is found that the loss of capacity is minor in typical cases.

The previous designs for distributed control of TCLs ensure that the indoor temperature remains within a pre-specified bound, but other QoS metrics, especially the frequency of turning on and off was not limited. In [19], presented at ACM BuildSys 2018, we propose a more advanced control architecture that reduces the

cycling rate of TCLs. We show through simulations that the proposed controller is able to reduce the cycling of individual TCLs compared to the previous designs with little loss in tracking of the grid-supplied reference signal.

## 6.28. Optimal control of energy storage

Energy storage revenue estimation is essential for analyzing financial feasibility of investment in batteries. In [22], we quantify the cycles of operation considering depth-of-discharge (DoD) of operational cycles and provide an algorithm to calculate equivalent 100% DoD cycles. This facilitates in comparing cycles of different DoDs. The battery life is frequently defined as a combination of cycle and calendar life. We propose a battery capacity degradation model based on the cycle and the calendar life and operational cycles. Using equivalent 100% DoD cycles and revenue generated, we calculate the dollars per cycle revenue of storage performing electricity price based arbitrage and ancillary services for load balancing in real time. Using PJM's (a regional transmission organization in the United States) real data we calculate short term and long term financial potential for the year of 2017. We observe that participating in ancillary services is significantly more beneficial for storage owners compared to participating in energy arbitrage.

Battery life is often described a combination of cycle life and calendar life. In [21], we propose a mechanism to limit the number of cycles of operation over a time horizon in an optimal arbitrage algorithm proposed in our previous work. The cycles of operation have to be tuned based on price volatility to maximize the battery life and arbitrage gains.

In [23], we analyze the effect of real time electricity price (RTP) on the amount of ancillary services required for load balancing in presence of responsive users, information asymmetry and forecast errors in demand and renewable energy sources (RES) generation. We consider a RTP that is determined by the forecasted generation and ramping cost. A community choice aggregator manages the load of all the consumers by setting the price. The consumer's objective is to minimize their overall cost of consumption. Ancillary services are called upon to balance the load in real time. With zero RES in the power network and a high degree of load flexibility, the proposed RTP flattens and the volatility in demand vanishes. However, in presence of RES the volatility in price and demand is reduced up to an extent and ancillary services are required for load balancing. The amount of ancillary services required increases with forecast errors. We also propose a real time algorithm that approximates the optimal consumer behavior under the complete information setting. Extensive numerical simulations are provided using real data from Pecan Street and Elia Belgium.

## 6.29. Dynamic matching models

The model of First Come First Served infinite bipartite matching was introduced in Caldentey, Kaplan and Weiss, 2009. In this model, there is a sequence of items that are chosen i.i.d. from a finite set $\mathcal{C}$ and an independent sequence of items that are chosen i.i.d. from a finite set $\mathcal{S}$, and a bipartite compatibility graph G between $\mathcal{C}$ and $\mathcal{S}$. Items of the two sequences are matched according to the compatibility graph, and the matching is FCFS, meaning that each item in the one sequence is matched to the earliest compatible unmatched item in the other sequence. In Adan and Weiss, 2012, a Markov chain associated with the matching was analyzed, a condition for stability was derived, and a product form stationary distribution was obtained. In [2], we present several new results that unveil the fundamental structure of the model. First, we provide a pathwise Loynes' type construction which enables to prove the existence of a unique matching for the model defined over all the integers. Second, we prove that the model is dynamically reversible: we define an exchange transformation in which we interchange the positions of each matched pair, and show that the items in the resulting permuted sequences are again independent and i.i.d., and the matching between them is FCFS in reversed time. Third, we obtain product form stationary distributions of several new Markov chains associated with the model. As a by-product, we compute useful performance measures, for instance the link lengths between matched items.

In [51], we propose an explicit construction of the stationary state of Extended Bipartite Matching (EBM) models, as defined in (Busic et. al., 2013). We use a Loynes-type backwards scheme similar in flavor to that in (Moyal et al., 2017), allowing to show the existence and uniqueness of a bi-infinite perfect matching under various conditions, for a large class of matching policies and of bipartite matching structures. The key algebraic element of our construction is the sub-additivity of a suitable stochastic recursive representation of the model, satisfied under most usual matching policies. By doing so, we also derive stability conditions for the system under general stationary ergodic assumptions, subsuming the classical markovian settings.

In [42], we consider holding costs for the items that are waiting to be matched. We model this problem as an MDP (Markov decision process) and study the discounted cost and the average cost case. We first consider a model with two types of supply and two types of demand items with an N matching graph. For linear cost function, we prove that an optimal matching policy gives priority to the end edges of the matching graph and is of threshold type for the diagonal edge. In addition, for the average cost problem, we compute the optimal threshold value. According to our preliminary numerical experiments, threshold-type policies performs also very well for more general bipartite graphs.

# 7. Bilateral Contracts and Grants with Industry

## 7.1. CRE with Huawei

18-month contract titled “Mathematical Modeling of 5G Ultra Dense Wireless Networks” between Inria represented by B. Blaszczyszyn (PI) and F. Baccelli, and Huawei comes to an end in December 2018. It aimed at investigating obstacle-based shadowing fields in the spatial models of cellular networks and efficient scheduling policies. Paul Keeler was hired by Inria as a research engineer thanks to this contract. The publication [39] is one of the deliverable of this contract.

## 7.2. CIFRE with Nokia

Contract with Nokia started in 2015 for the co-advising by B. Blaszczyszyn of a PhD student of Nokia, Dalia-Georgiana Herculea came to an end in December 2018. Dalia-Georgiana Herculea has successfully defended her PhD Thesis in November 2018.

## 7.3. CIFRE with Orange

Contract with Orange started in 2017 and continued in 2018 for the co-advising by B. Blaszczyszyn of a PhD student of Orange, Quentin Le Gall.

# 8. Partnerships and Cooperations

## 8.1. Regional Initiatives

### *8.1.1. Laboratory of Information, Networking and Communication Sciences (LINCS)*

Dyogene participates in LINCS https://www.lincs.fr/, a research centre co-founded by Inria, Institut Mines-Télécom, UPMC and Alcatel-Lucent Bell Labs (currently Nokia Bell Labs) dedicated to research and innovation in the domains of future information and communication networks, systems and services. S. Meyn [Unversity of Florida] was invited professor by LINCS and ENS from July to December 2018.

### *8.1.2. PGMO*

Dyogene participates to the PGMO (Gaspard Monge Program for Optimization, operations research, and their interactions with data science) via the project a 2 year project “Distributed control of flexible loads” funded through the ICODE/IROE call. This is a collaborative project between University Paris-Sud (PI: Gilles Stoltz) and Inria (PI: Ana Busic). Post-doc Cheng Wan was financed by this project from Feb-Nov 2018.

## 8.2. National Initiatives

### *8.2.1. GdR GeoSto*

Members of Dyogene participate in Research Group GeoSto (Groupement de recherche, GdR 3477) http://gdr-geostoch.math.cnrs.fr/ on Stochastic Geometry led by and David Coupier [Université de Valenciennes].

This is a collaboration framework for all French research teams working in the domain of spatial stochastic modeling, both on theory development and in applications. This year DYOGENE has co-organized yearly conference of the GdR *Stochastic Geometry Days 2018* 14–18 mai 2018 Paris (France); https://geosto-2018.sciencesconf.org/.

### *8.2.2. GdR RO*

Members of Dyogene participate in GdR-RO (Recherche Opérationelle; GdR CNRS 3002), http://gdrro.lip6.fr/, working group COSMOS (Stochastic optimization and control, modeling and simulation), lead by A. Busic and E. Hyon (LIP 6); http://gdrro.lip6.fr/?q=node/78

### *8.2.3. ANR JCJC PARI*

Probabilistic Approach for Renewable Energy Integration: Virtual Storage from Flexible Loads. The project started in January 2017. PI — A. Bušić. This project is motivated by current and projected needs of a power grid with significant renewable energy integration. Renewable energy sources such as wind and solar have a high degree of unpredictability and time variation, which makes balancing demand and supply challenging. There is an increased need for ancillary services to smooth the volatility of renewable power. In the absence of large, expensive batteries, we may have to increase our inventory of responsive fossil-fuel generators, negating the environmental benefits of renewable energy. The proposed approach addresses this challenge by harnessing the inherent flexibility in demand of many types of loads. The objective of the project is to develop decentralized control for automated demand dispatch, that can be used by grid operators as ancillary service to regulate demand-supply balance at low cost. We call the resource obtained from these techniques virtual energy storage (VES). Our goal is to create the necessary ancillary services for the grid that are environmentally friendly, that have low cost and that do not impact the quality of service (QoS) for the consumers. Besides respecting the needs of the loads, the aim of the project is to design local control solutions that require minimal communications from the loads to the centralized entity. This is possible through a systems architecture that includes the following elements: i) local control at each load based on local measurements combined with a grid-level signal; ii) frequency decomposition of the regulation signal based on QoS and physical constraints for each class of loads.

## 8.3. International Initiatives

### *8.3.1. Inria International Partners*

#### *8.3.1.1. IFCAM Project "Geometric statistics of stationary point processes"*

B. Błaszczyszyn and Yogeshwaran D. from Indian Statistical Institute (ISI), Bangalore, have got in 2018 the approval from Indo-French Centre for Applied Mathematics (IFCAM), for their joint project on "Geometric statistics of stationary point processes" for the period 2018–2021. B. Błaszczyszyn was visiting ISI Bangalore for two weeks in November–December 2018.

#### *8.3.1.2. Informal International Partners*

- University of Florida: collaborations with Prof Sean Meyn (ECE), Associate Prof Prabir Barooah (MAE), and the PhD students: A. Devraj (ECE), A. Coffman (MAE), N. Cammardella (ECE), J. Mathias (ECE).

## 8.4. International Research Visitors

### *8.4.1. Visits of International Scientists*

- D. Yogeshwaran [Indian Statistical Institute, Bangalore, India]
- S. Meyn [University of Florida, USA] was invited Prof at ENS and LINCS, July - December 2018

*8.4.1.1. Internships*

- Master Probabilités et Modèles aléatoires UPMC, Walid Ghanem, *Hydrodynamic limit of a network with moving servers*, 04-07/2018, encadrant Christine Fricker.
- Master MASH (Mathématiques appliquées aux sciences humaines) ENS-Paris Dauphine University, *Using customer oriented policies based on probabilistic methods to enhance the Bike Sharing System Velib'*, 08-011/2018, encadrants Christine Fricker et Laurent Massoulié.
- Akshay Goel [Kyushu University, Fukuoka, Japan] Mars 2018,
- Tokuyama Kiichi [Tokyo Tech, Tokyo, Japan], April 2018,

### *8.4.2. Visits to International Teams*

*8.4.2.1. Research Stays Abroad*

- B. Błaszczyszyn was visiting Yogeshwaran D. at the Indian Statistical Institute Bangalore for two weeks in November–December 2018 (IFCAM project).
- A. Busic was a long-term participant (March-Mai 2018) of the Real-Time Decision Making program, Simons Institute, UC Berkeley, USA; https://simons.berkeley.edu/programs/realtime2018

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Organisation*

*9.1.1.1. Member of the Organizing Committees*

- B. Błaszczyszyn was in the Organizing Commitee of *Stochastic Geometry Days 2018* 14–18 mai 2018 Paris (France); https://geosto-2018.sciencesconf.org/.
- A. Busic was in the Organizing Commitee of *ALEA Days 2018* 12–16 March 2018 at CIRM, Luminy (France); https://conferences.cirm-math.fr/1776.html

### *9.1.2. Scientific Events Selection*

*9.1.2.1. Member of the Editorial Boards*

- M. Lelarge: IEEE's Transactions on Network Science and Engineering, Bernoulli Journal, Queueing Systems.

### *9.1.3. Invited Talks*

- Analysis of Algorithms 2018, Uppsala (Sweden), June 2018.
- European Conference on Queueing Theory (ECQT) 2018, Jerusalem, July 2018.
- Rencontres de Probabilités de Rouen 2018, September 2018, invited talk.
- Societal Networks, RTDM program, Simons Institute, UC Berkeley, USA, March 2018, invited talk; video: https://simons.berkeley.edu/talks/ana-busic-3-29-18.
- Advances in Modelling and Control for Power Systems of the Future (CAESARS 2018), EDF Palaiseau, September 2018, invited talk.

## 9.2. Teaching - Supervision - Juries

### *9.2.1. Teaching*

- Licence: B. Blaszczyszyn (Cours) **Théorie de l'information et du codage** 24 heqTD, L3, ENS, France.
- Licence: A. Busic (Cours) and S. Samain (TD) Structures et algorithmes aléatoires 60heqTD, L3, ENS, France.
- Master: B. Blaszczyszyn (Cours) **Processus ponctuels, graphes aléatoires et géeométrie stochastique** 39heqTD, M2 Probabilités et Modèles Aléatoires, UPMC, France
- Master: A. Busic (Cours) and L. Stephan (TD) **Modéles et algorithmes de réseaux** 60heqTD, M1, ENS, France.
- Master: A. Busic (Cours) **Fondements de la modélisation des réseaux** 18 heqTD, M2 MPRI, France.
- Master: M. Lelarge (Cours) **Deep Learning Do it Yourself**, M1, ENS
- Doctorat: A. Busic (Cours) **Markov chains and exact sampling**, 7heqTD, Ecole thématique CNRS, MathExp 2018 "Mathématiques expérimentales: méthodes et pratiques", 21 mai-1 juin 2018, Saint Flour, France. https://mathexp2018.sciencesconf.org/

### *9.2.2. Supervision*

- PhD: Dalia-Georgiana Herculea "Les Hyperfractales pour la Modelisation des Reseaux sans Fil" since October 2016, defence 21 November 2018; PhD CIFRE co-advised by B. Błaszczyszyn, and Ph. Jacquet.
- PhD: Alexandre Hollocou, defense December 19 2018, Nouvelles approches pour le partitionnement de graphes, co-advised by M. Lelarge and T. Bonald (Telecom ParisTech)
- PhD in progress: Léo Miolane, since 2016, High dimensional statistics, advised by M. Lelarge
- PhD in progress: Alexis Galland, since 2017, Deep Learning on Graphs, advised by M. Lelarge
- PhD in progress: Quentin Le Gall "Crowd networking : modélisation de la connectivité D2D" since October 2017; PhD CIFRE co-advised by B. Błaszczyszyn and E. Cali (Orange).
- PhD in progress: Antoine Brochard "Signal processing for point processes and statistical learning for telecommunications", since September 2018; PhD CIFRE co-advised by B. Błaszczyszyn and Georgios Paschos (Huawei).
- PhD in progress: Md Umar Hashmi, Decentralized control for renewable integration in smartgrids, since from December 2015, advised by A. Busic.
- PhD in progress: Sébastien Samain, Monte Carlo methods for performance evaluation and reinforcement learning, since November 2016, advised by A. Busic.
- PhD in progress: Arnaud Cadas, Dynamic matching models, since October 2017, supervised by A. Busic.

### *9.2.3. Juries*

- B. Błaszczyszyn, member of the PhD defense jury of Dalia-Herculea Popescu; 21 November 2018.
- A. Busic, member of the PhD defense jury of J. Horta (Télécom ParisTech, France); 16 Avril 2018.
- C. Fricker, PhD defense of Farah Slim (Orange- Bretagne Loire University), 13/03/2018.

## 9.3. Popularization

### *9.3.1. Internal or external Inria responsibilities*

- B. Błaszczyszyn is an ENS adjunct professor since September 2018.
- B. Błaszczyszyn and A. Busic are members of the ENS Computer Science Department Board (Conseil du Laboratoire).
- A. Busic is member of the Committee for the Technological Development, Inria Paris

### *9.3.2. Education*

C. Fricker: member of the jury of *Agrégation de Mathématiques*.

# 10. Bibliography

## Publications of the year

### Articles in International Peer-Reviewed Journal

[1] E. ABBE, L. MASSOULIÉ, A. MONTANARI, A. SLY, N. SRIVASTAVA.*Group synchronization on grids*, in "Mathematical Statistics and Learning", September 2018, https://arxiv.org/abs/1706.08561 , https://hal.archives-ouvertes.fr/hal-01940467

[2] I. ADAN, A. BUŠIC, J. MAIRESSE, G. WEISS.*Reversibility and further properties of FCFS infinite bipartite matching*, in "Mathematics of Operations Research", May 2018, vol. 43, n$^{o}$ 2, p. 347-692, https://arxiv.org/abs/1507.05939 - 27 pages, 12 figures [*DOI :* 10.1287/MOOR.2017.0874], https://hal.inria.fr/hal-01273897

[3] J. BARRE, M. LELARGE, J. MITSCHE.*On rigidity, orientability, and cores of random graphs with sliders*, in "Random Structures & Algorithms", May 2018, vol. 52, n$^{o}$ 3, p. 419-453, https://hal.archives-ouvertes.fr/hal-01963872

[4] C. BORDENAVE, M. LELARGE, L. MASSOULIÉ.*Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs*, in "Annals of Probability", January 2018, vol. 46, n$^{o}$ 1, p. 1-71, https://hal.archives-ouvertes.fr/hal-01940486

[5] A. BUŠIC, S. MEYN.*Ordinary Differential Equation Methods for Markov Decision Processes and Application to Kullback–Leibler Control Cost*, in "SIAM Journal on Control and Optimization", February 2018, vol. 56, n$^{o}$ 1, p. 343-366, https://arxiv.org/abs/1605.04591 [*DOI :* 10.1137/16M1100204], https://hal.archives-ouvertes.fr/hal-01968522

[6] F. CALTAGIRONE, M. LELARGE, L. MIOLANE.*Recovering Asymmetric Communities in the Stochastic Block Model*, in "IEEE Transactions on Network Science and Engineering", July 2018, vol. 5, n$^{o}$ 3, p. 237-246, https://hal.archives-ouvertes.fr/hal-01963866

[7] A. CHATTOPADHYAY, B. BLASZCZYSZYN, E. ALTMAN.*Location Aware Opportunistic Bandwidth Sharing between Static and Mobile Users with Stochastic Learning in Cellular Networks*, in "IEEE Transactions on Mobile Computing", 2018, https://arxiv.org/abs/1608.04260 - 16 Pages, 1 Figure, 1 Table, https://hal.inria.fr/hal-01401007

[8] A. CHATTOPADHYAY, B. BLASZCZYSZYN, E. ALTMAN.*Two-tier cellular networks for throughput maximization of static and mobile users*, in "IEEE Transactions on Wireless Communications", 2018, https://arxiv.org/abs/1605.07341 - title of the previous version: "Cell planning for mobility management in heterogeneous cellular networks" [*DOI :* 10.1109/TWC.2018.2887386], https://hal.inria.fr/hal-01331936

[9] A. CHATTOPADHYAY, B. BLASZCZYSZYN, H. P. KEELER.*Gibbsian On-Line Distributed Content Caching Strategy for Cellular Networks*, in "IEEE Transactions on Wireless Communications", February 2018, vol. 17, n$^{o}$ 2, p. 969 - 981, https://arxiv.org/abs/1610.02318 [*DOI :* 10.1109/TWC.2017.2772911], https://hal.inria.fr/hal-01401010

[10] Y. CHEN, A. BUŠIC, S. MEYN.*Ergodic Theory for Controlled Markov Chains with Stationary Inputs*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", 2018, vol. 28, n$^{o}$ 1, p. 79-111, https://hal.archives-ouvertes.fr/hal-01672476

[11] Y. CHEN, A. BUŠIC, S. MEYN.*Estimation and Control of Quality of Service in Demand Dispatch*, in "IEEE Transactions on Smart Grid", 2018, vol. 9, n$^{o}$ 5, p. 5348 - 5356, https://hal.archives-ouvertes.fr/hal-01672458

[12] L. GULIKERS, M. LELARGE, L. MASSOULIÉ.*An impossibility result for reconstruction in the degree-corrected stochastic block model*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", October 2018, vol. 28, n$^{o}$ 5, p. 3002-3027, https://arxiv.org/abs/1511.00546 , https://hal.archives-ouvertes.fr/hal-01940494

[13] 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

[14] L. MASSOULIÉ, K. XU.*On the Capacity of Information Processing Systems*, in "Operations Research", April 2018, vol. 66, n$^{o}$ 2, p. 568-586, https://hal.archives-ouvertes.fr/hal-01940447

### International Conferences with Proceedings

[15] A. BOUILLARD, F. MATHIEU, P. SEHIER, T. DEISS.*Dimensionnement du fronthaul 5G : c'est simple comme un coût de file*, 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.inria.fr/hal-01787187

[16] A. BUŠIC, S. MEYN.*Action-Constrained Markov Decision Processes With Kullback-Leibler Cost*, in "Conference On Learning Theory (COLT)", Stockholm, Sweden, July 2018, https://arxiv.org/abs/1807.10244 , https://hal.archives-ouvertes.fr/hal-01968536

[17] N. CAMMARDELLA, J. MATHIAS, M. KIENER, A. BUŠIC, S. MEYN.*Balancing California's Grid Without Batteries*, in "57th IEEE Conference on Decision and Control (CDC 2018)", Miami, United States, December 2018, https://hal.archives-ouvertes.fr/hal-01968606

[18] A. R. COFFMAN, A. BUŠIC, P. BAROOAH.*A Study of Virtual Energy Storage From Thermostatically Controlled Loads Under Time-Varying Weather Conditions*, in "International High Performance Buildings Conference", Purdue, United States, July 2018, https://hal.archives-ouvertes.fr/hal-01968587

[19] A. R. COFFMAN, A. BUŠIC, P. BAROOAH.*Virtual energy storage from TCLs using QoS preserving local randomized control*, in "5th Conference on Systems for Built Environments (BuildSys '18)", Shenzen, China, November 2018, p. 93-102, https://hal.archives-ouvertes.fr/hal-01968602

[20] P. S. DESTER, C. FRICKER, H. MOHAMED.*Stationary Distribution Analysis of a Queueing Model with Local Choice*, in "29th International Conference on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms (AofA 2018)", Uppsala, Sweden, J. A. FILL, M. D. WARD (editors), LIPIcs-Leibniz International Proceedings in Informatics, June 2018, vol. 110, n$^{o}$ 22 [*DOI :* 10.4230/LIPICS.AOFA.2018.22], https://hal.inria.fr/hal-01666326

[21] M. U. HASHMI, A. BUŠIC.*Limiting Energy Storage Cycles of Operation*, in "10th IEEE Green Technologies Conference (GreenTech 2018)", Austin, TX, United States, April 2018 [*DOI :* 10.1109/GREENTECH.2018.00022], https://hal.archives-ouvertes.fr/hal-01806487

[22] M. U. HASHMI, W. LABIDI, A. BUŠIC, S.-E. ELAYOUBI, T. CHAHED.*Long-Term Revenue Estimation for Battery Performing Arbitrage and Ancillary Services*, in "SmartGridComm 2018 - IEEE International Conference Communications, Control, and Computing Technologies for Smart Grids", Aalborg, Denmark, October 2018, p. 1-7, https://hal.archives-ouvertes.fr/hal-01867341

[23] M. U. HASHMI, D. MUTHIRAYAN, A. BUŠIC.*Effect of Real-Time Electricity Pricing on Ancillary Service Requirements*, in "Ninth International Conference on Future Energy Systems (ACM e-Energy 2018). Workshops", Karlsruhe, Germany, June 2018, p. 550-555 [*DOI :* 10.1145/3208903.3214350], https://hal.archives-ouvertes.fr/hal-01968604

[24] P. JACQUOT, C. WAN.*Routing Game on Parallel Networks: the Convergence of Atomic to Nonatomic*, in "CDC 2018 - IEEE 57th Conference on Decision and Control", Miami, United States, Proceedings of the 57th IEEE Conference on Decision and Control, IEEE, December 2018, vol. 1, https://hal.archives-ouvertes.fr/hal-01762547

[25] D. POPESCU, P. JACQUET.*Vehicle-to-Infrastructure Communications Design in Urban Hyperfractals*, in "19th IEEE International Workshop on Signal Processing Advances in Wireless Communications, SPAWC 2018", Kalamata, Greece, June 2018, https://hal.inria.fr/hal-01787602

### Conferences without Proceedings

[26] T. BONALD, B. CHARPENTIER, A. GALLAND, A. HOLLOCOU.*Hierarchical Graph Clustering using Node Pair Sampling*, in "MLG 2018 - 14th International Workshop on Mining and Learning with Graphs", London, United Kingdom, August 2018, https://hal.archives-ouvertes.fr/hal-01887669

[27] T. BONALD, A. HOLLOCOU, M. LELARGE.*Weighted Spectral Embedding of Graphs*, in "56th Annual Allerton Conference on Communication, Control, and Computing", Urbana-Champaign, United States, October 2018, https://hal.archives-ouvertes.fr/hal-01887680

[28] A. MUKHOPADHYAY, N. HEGDE, M. LELARGE.*Optimal Content Replication and Request Matching in Large Caching Systems*, in "INFOCOM 2018 - IEEE International Conference on Computer Communications", Honolulu, United States, April 2018, https://arxiv.org/abs/1801.02889 , https://hal.archives-ouvertes.fr/hal-01963864

[29] K. SCAMAN, F. BACH, S. BUBECK, Y. T. LEE, L. MASSOULIÉ.*Optimal Algorithms for Non-Smooth Distributed Optimization in Networks*, in "Advances In Neural Information Processing systems", Montreal, Canada, December 2018, https://arxiv.org/abs/1806.00291 - 17 pages, https://hal.archives-ouvertes.fr/hal-01957013

### Scientific Books (or Scientific Book chapters)

[30] B. BLASZCZYSZYN, M. HAENGGI, H. P. KEELER, S. MUKHERJEE.*Stochastic Geometry Analysis of Cellular Networks*, Cambridge University Press, March 2018 [*DOI :* 10.1017/9781316677339], https://hal.inria.fr/hal-01958627

[31] Y. CHEN, M. U. HASHMI, J. MATHIAS, A. BUŠIC, S. MEYN.*Distributed Control Design for Balancing the Grid Using Flexible Loads*, in "IMA Volume on the Control of Energy Markets and Grids", S. MEYN, T. SAMAD, I. HISKENS, J. STOUSTRUP (editors), Springer, 2018, vol. 162, p. 383-411, https://hal.archives-ouvertes.fr/hal-01656726

### Scientific Popularization

[32] P. JACQUET, D. POPESCU, B. MANS.*Broadcast Speedup in Vehicular Networks via Information Teleportation*, in "The 43rd Annual IEEE Conference on Local Computer Networks (LCN 2018)", Chicago, United States, October 2018, https://hal.inria.fr/hal-01963430

[33] P. JACQUET, D. POPESCU, B. MANS.*Energy Trade-offs for end-to-end Communications in Urban Vehicular Networks exploiting an Hyperfractal Model*, in "MSWIM DIVANet", Montreal, Canada, IEEE, October 2018, https://hal.inria.fr/hal-01674685

## Patents and standards

[34] F. MATHIEU, A. BOUILLARD.*A Method for Scoring Objects, a Method of Ranking an Object, a Software Product, and an Apparatus*, February 2018, n$^{o}$ 18305203.4, https://hal.archives-ouvertes.fr/hal-01916427

[35] S. MEYN, A. BUŠIC.*Using loads with discrete finite states of power to provide ancillary services for a power grid*, July 2018, n$^{o}$ 15740328, https://hal.archives-ouvertes.fr/hal-01968598

### Other Publications

[36] F. BACCELLI, M.-O. HAJI-MIRSADEGHI, A. KHEZELI.*On the Dimension of Unimodular Discrete Spaces, Part I: Definitions and Basic Properties*, January 2019, https://arxiv.org/abs/1807.02980 - working paper or preprint, https://hal.inria.fr/hal-01976265

[37] F. BACCELLI, M.-O. HAJI-MIRSADEGHI, A. KHEZELI.*On the Dimension of Unimodular Discrete Spaces, Part II: Relations with Growth Rate*, January 2019, https://arxiv.org/abs/1808.02551 - working paper or preprint, https://hal.inria.fr/hal-01976257

[38] B. BLASZCZYSZYN, M. K. KARRAY.*Performance analysis of cellular networks with opportunistic scheduling using queueing theory and stochastic geometry*, June 2018, working paper or preprint, https://hal.inria.fr/hal-01824986

[39] B. BLASZCZYSZYN, H. P. KEELER.*Determinantal thinning of point processes with network learning applications*, December 2018, https://arxiv.org/abs/1810.08672 - 8 pages, 6 figures. All code available online: https://github.com/hpaulkeeler/DetPoisson_MATLAB [*DOI :* 10.08672], https://hal.inria.fr/hal-01958924

[40] B. BLASZCZYSZYN, P. MUHLETHALER.*Analyzing LoRa long-range, low-power, wide-area networks using stochastic geometry*, December 2018, https://arxiv.org/abs/1811.01886 - 8 pages, https://hal.inria.fr/hal-01958939

[41] A. BROCHARD, B. BLASZCZYSZYN, S. MALLAT, S. ZHANG.*Statistical learning of geometric characteristics of wireless networks*, 2018, https://arxiv.org/abs/1812.08265 - working paper or preprint, https://hal.inria.fr/hal-01962191

[42] A. CADAS, A. BUŠIC, J. DONCEL.*Optimal Control of Dynamic Bipartite Matching Models*, 2018, https://arxiv.org/abs/1810.08541 - working paper or preprint [*DOI :* 10.08541], https://hal.archives-ouvertes.fr/hal-01968549

[43] A. M. DEVRAJ, A. BUŠIC, S. MEYN.*Zap Meets Momentum: Stochastic Approximation Algorithms with Optimal Convergence Rate*, September 2018, https://arxiv.org/abs/1809.06277 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01968558

[44] M. U. HASHMI.*Load Flexibility for Price based Demand Response*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01911254

[45] H. HENDRIKX, F. BACH, L. MASSOULIÉ.*Accelerated decentralized optimization with local updates for smooth and strongly convex objectives*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01893568

[46] A. JAGANNATH, P. LOPATTO, L. MIOLANE.*Statistical thresholds for Tensor PCA*, December 2018, https://arxiv.org/abs/1812.03403 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01960964

[47] H. P. KEELER, B. JAHNEL, O. MAYE, D. ASCHENBACH, M. BRZOZOWSKI.*Disruptive events in high-density cellular networks*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01960597

[48] Q. LE GALL, B. BLASZCZYSZYN, E. CALI, T. EN-NAJJARY.*The Influence of Canyon Shadowing on Device-to-Device Connectivity in Urban Scenario*, October 2018, https://arxiv.org/abs/1810.03498 - 8 pages, 7 figures, 2 tables [*DOI :* 10.03498], https://hal.inria.fr/hal-01891024

[49] L. MIOLANE.*Phase transitions in spiked matrix estimation: information-theoretic analysis*, December 2018, https://arxiv.org/abs/1806.04343 - These notes present in a unified manner recent results (as well as new developments) on the information-theoretic limits in spiked matrix/tensor estimation, https://hal.archives-ouvertes.fr/hal-01960925

[50] L. MIOLANE, A. MONTANARI.*The distribution of the Lasso: Uniform control over sparse balls and adaptive parameter tuning*, December 2018, https://arxiv.org/abs/1811.01212 - 68 pages, 2 figures, https://hal.archives-ouvertes.fr/hal-01960945

[51] P. MOYAL, A. BUŠIC, J. MAIRESSE.*Loynes construction for the extended bipartite matching*, March 2018, https://arxiv.org/abs/1803.02788 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01968565

# Project-Team EVA

## Wireless Networking for Evolving & Adaptive Applications

RESEARCH CENTER
**Paris**

THEME
**Networks and Telecommunications**

# Table of contents

# Project-Team EVA

*Creation of the Team: 2015 April 01, updated into Project-Team: 2016 May 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.2. - Networks
A1.2.1. - Dynamic reconfiguration
A1.2.2. - Supervision
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
A1.2.8. - Network security
A1.2.9. - Social Networks
A1.4. - Ubiquitous Systems
A1.6. - Green Computing
A2.3. - Embedded and cyber-physical systems
A2.3.1. - Embedded systems
A2.3.2. - Cyber-physical systems
A2.3.3. - Real-time systems
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A4. - Security and privacy
A4.1. - Threat analysis
A4.1.1. - Malware analysis
A4.1.2. - Hardware attacks
A4.4. - Security of equipment and software
A4.5. - Formal methods for security
A4.6. - Authentication
A4.7. - Access control
A5.10. - Robotics
A5.10.6. - Swarm robotics
A5.10.8. - Cognitive robotics and systems
A6. - Modeling, simulation and control
A9. - Artificial intelligence
A9.2. - Machine learning
A9.7. - AI algorithmics

**Other Research Topics and Application Domains:**

B5.1. - Factory of the future
B6. - IT and telecom
B6.2. - Network technologies
B6.2.1. - Wired technologies
B6.2.2. - Radio technology
B6.3.2. - Network protocols
B6.3.3. - Network Management
B6.3.4. - Social Networks
B6.4. - Internet of things
B6.6. - Embedded systems
B7. - Transport and logistics
B7.1.1. - Pedestrian traffic and crowds
B7.1.2. - Road traffic
B7.2. - Smart travel
B7.2.1. - Smart vehicles
B7.2.2. - Smart road
B8. - Smart Cities and Territories
B8.1. - Smart building/home
B8.1.1. - Energy for smart buildings
B8.1.2. - Sensor networks for smart buildings
B8.2. - Connected city
B8.4. - Security and personal assistance
B8.4.1. - Crisis management

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Paul Muhlethaler [Team leader, Inria, Senior Researcher, HDR]
Pascale Minet [Inria, Researcher, HDR]
Malisa Vucinic [Inria, from Nov 2018, Starting Research position]
Thomas Watteyne [Inria, Advanced Research position]

**External Collaborators**

Nadjib Achir [Univ Paris-Nord, HDR]
Saadi Boudjit [Univ Paris-Nord]
Selma Boumerdassi [CNAM]
Samia Bouzefrane [CNAM]
Philippe Jacquet [Bell Labs (Nokia), HDR]
Anis Laouiti [Telecom Sud-Paris, HDR]
Dana Marinca [Univ de Versailles Saint-Quentin-en-Yvelines]
Eric Renault [Telecom Sud-Paris, HDR]

**Technical Staff**

Remy Leone [Inria, until Jan 2018]
Yasuyuki Tanaka [Inria]
Trifun Savic [Inria, GeoBot project, from February 2019]

**PhD Students**

Amar Abane [CNAM, from Mar 2018]
Keoma Brun-Laguna [Inria]
Iman Hmedoush [Inria, from Oct 2018]
Jonathan Munoz Soto [Inria]
Mamadou Sangare [Univ of Conakry]
Abdallah Sobehy [Univ Paris-Saclay]

**Post-Doctoral Fellow**
Tengfei Chang [Inria]

**Administrative Assistant**
Andree Nicole Chamroux [Inria, from Apr 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

It is forecast that the vast majority of Internet connections will be wireless. The EVA project grasps this opportunity and focuses on wireless communication. EVA tackles challenges related to providing efficient communication in wireless networks and, more generally, in all networks that are not already organized when set up, and consequently need to evolve and spontaneously find a match between application requirements and the environment. These networks can use opportunistic and/or collaborative communication schemes. They can evolve through optimization and self-learning techniques. Every effort is made to ensure that the results provided by EVA have the greatest possible impact through standardization. The miniaturization and ubiquitous nature of computing devices has opened the way to the deployment of a new generation of wireless (sensor) networks. These networks are central to the work in EVA, as EVA focuses on such crucial issues as power conservation, connectivity, determinism, reliability and latency. Wireless Sensor Network (WSN) deployments are also be a new key subject, especially for emergency situations (e.g. after a disaster). Industrial process automation and environmental monitoring are considered in greater depth.

# 3. Research Program

## 3.1. Pitch

**Designing Tomorrow's Internet of (Important) Things**

Inria-EVA is a leading research team in low-power wireless communications. The team pushes the limits of low-power wireless mesh networking by applying them to critical applications such as industrial control loops, with harsh reliability, scalability, security and energy constraints. Grounded in real-world use cases and experimentation, EVA co-chairs the IETF 6TiSCH standardization working group, co-leads Berkeley's OpenWSN project and works extensively with Analog Devices' SmartMesh IP networks. Inria-EVA is the birthplace of the Wattson Elements startup and the Falco solution. The team is associated with Prof. Glaser's (UC Berkeley) and Prof. Kerkez (U. Michigan) through the REALMS associate research team, and with OpenMote through a long-standing Memorandum of Understanding.

## 3.2. Physical Layer

We study how advanced physical layers can be used in low-power wireless networks. For instance, collaborative techniques such as multiple antennas (e.g. Massive MIMO technology) can improve communication efficiency. The core idea is to use massive network densification by drastically increasing the number of sensors in a given area in a Time Division Duplex (TDD) mode with time reversal. The first period allows the sensors to estimate the channel state and, after time reversal, the second period is to transmit the data sensed. Other techniques, such as interference cancellation, are also possible.

## 3.3. Wireless Access

Medium sharing in wireless systems has received substantial attention throughout the last decade. HiPERCOM2 has provided models to compare TDMA and CSMA. HiPERCOM2 has also studied how network nodes must be positioned to optimize the global throughput.

EVA pursues modeling tasks to compare access protocols, including multi-carrier access, adaptive CSMA (particularly in VANETs), as well as directional and multiple antennas. There is a strong need for determinism in industrial networks. The EVA team focuses particularly on scheduled medium access in the context of deterministic industrial networks; this involves optimizing the joint time slot and channel assignment. Distributed approaches are considered, and the EVA team determines their limits in terms of reliability, latency and throughput. Furthermore, adaptivity to application or environment changes are taken into account.

## 3.4. Coexistence of Wireless Technologies

Wireless technologies such as cellular, low-power mesh networks, (Low-Power) WiFi, and Bluetooth (low-energy) can reasonably claim to fit the requirements of the IoT. Each, however, uses different trade-offs between reliability, energy consumption and throughput. The EVA team will study the limits of each technology, and will develop clear criteria to evaluate which technology is best suited to a particular set of constraints.

Coexistence between these different technologies (or different deployments of the same technology in a common radio space) is a valid point of concern.

The EVA team aims at studying such coexistence, and, where necessary, propose techniques to improve it. Where applicable, the techniques will be put forward for standardization. Multiple technologies can also function in a symbiotic way.

For example, to improve the quality of experience provided to end users, a wireless mesh network can transport sensor and actuator data in place of a cellular network, when and where cellular connectivity is poor.

The EVA team will study how and when different technologies can complement one another. A specific example of a collaborative approach is Cognitive Radio Sensor Networks (CRSN).

## 3.5. Energy-Efficiency and Determinism

Reducing the energy consumption of low-power wireless devices remains a challenging task. The overall energy budget of a system can be reduced by using less power-hungry chips, and significant research is being done in that direction. That being said, power consumption is mostly influenced by the algorithms and protocols used in low-power wireless devices, since they influence the duty-cycle of the radio.

EVA will search for energy-efficient mechanisms in low-power wireless networks. One new requirement concerns the ability to predict energy consumption with a high degree of accuracy. Scheduled communication, such as the one used in the IEEE 802.15.4 TSCH (Time Slotted CHannel Hopping) standard, and by IETF 6TiSCH, allows for a very accurate prediction of the energy consumption of a chip. Power conservation will be a key issue in EVA.

To tackle this issue and match link-layer resources to application needs, EVA's 5-year research program around Energy-Efficiency and Determinism centers around 3 studies:

- Performance Bounds of a TSCH network. We propose to study a low-power wireless TSCH network as a Networked Control System (NCS), and use results from the NCS literature. A large number of publications on NCS, although dealing with wireless systems, consider wireless links to have perfect reliability, and do not consider packet loss. Results from these papers can not therefore be applied directly to TSCH networks. Instead of following a purely mathematical approach to model the network, we propose to use a non-conventional approach and build an empirical model of a TSCH network.

- Distributed Scheduling in TSCH networks. Distributed scheduling is attractive due to its scalability and reactivity, but might result in a sub-optimal schedule. We continue this research by designing a distributed solution based on control theory, and verify how this solution can satisfy service level agreements in a dynamic environment.

## 3.6. Network Deployment

Since sensor networks are very often built to monitor geographical areas, sensor deployment is a key issue. The deployment of the network must ensure full/partial, permanent/intermittent coverage and connectivity. This technical issue leads to geometrical problems which are unusual in the networking domain.

We can identify two scenarios. In the first one, sensors are deployed over a given area to guarantee full coverage and connectivity, while minimizing the number of sensor nodes. In the second one, a network is re-deployed to improve its performance, possibly by increasing the number of points of interest covered, and by ensuring connectivity. EVA will investigate these two scenarios, as well as centralized and distributed approaches. The work starts with simple 2D models and will be enriched to take into account more realistic environment: obstacles, walls, 3D, fading.

## 3.7. Data Gathering and Dissemination

A large number of WSN applications mostly do data gathering (a.k.a "convergecast"). These applications usually require small delays for the data to reach the gateway node, requiring time consistency across gathered data. This time consistency is usually achieved by a short gathering period.

In many real WSN deployments, the channel used by the WSN usually encounters perturbations such as jamming, external interferences or noise caused by external sources (e.g. a polluting source such as a radar) or other coexisting wireless networks (e.g. WiFi, Bluetooth). Commercial sensor nodes can communicate on multiple frequencies as specified in the IEEE 802.15.4 standard. This reality has given birth to the multichannel communication paradigm in WSNs.

Multichannel WSNs significantly expand the capability of single-channel WSNs by allowing parallel transmissions, and avoiding congestion on channels or performance degradation caused by interfering devices.

In EVA, we will focus on raw data convergecast in multichannel low-power wireless networks. In this context, we are interested in centralized/distributed algorithms that jointly optimize the channel and time slot assignment used in a data gathering frame. The limits in terms of reliability, latency and bandwidth will be evaluated. Adaptivity to additional traffic demands will be improved.

## 3.8. Self-Learning Networks

To adapt to varying conditions in the environment and application requirements, the EVA team will investigate self-learning networks. Machine learning approaches, based on experts and forecasters, will be investigated to predict the quality of the wireless links in a WSN. This allows the routing protocol to avoid using links exhibiting poor quality and to change the route before a link failure. Additional applications include where to place the aggregation function in data gathering. In a content delivery network (CDN), it is very useful to predict the popularity, expressed by the number of solicitations per day, of a multimedia content. The most popular contents are cached near the end-users to maximize the hit ratio of end-users' requests. Thus the satisfaction degree of end-users is maximized and the network overhead is minimized.

## 3.9. Security Trade-off in Constrained Wireless Networks

Ensuring security is a sine qua non condition for the widespread acceptance and adoption of the IoT, in particular in industrial and military applications. While the Public-Key Infrastructure (PKI) approach is ubiquitous on the traditional Internet, constraints in terms of embedded memory, communication bandwidth and computational power make translating PKI to constrained networks non-trivial.

In the IETF 6TiSCH working group, and through the work on Malisa Vucinic as part of the H2020 ARMOUR project, we have started to work on a "Minimal Security" solution at the IETF. This solution is based on pre-shared keying material, and offers mutual authentication between each node in the network and central security authority, replay protection and key rotation.

# 4. Application Domains

## 4.1. Industrial Process Automation

Wireless networks have become ubiquitous and are an integral part of our daily lives. These networks are present in many application domains; the most important are detailed in this section.

Networks in industrial process automation typically perform **monitoring and control** tasks. Wired industrial communication networks, such as HART [0], have been around for decades and, being wired, are highly reliable. Network administrators tempted to "go wireless" expect the same reliability. Reliable process automation networks – especially when used for control – often impose stringent latency requirements. Deterministic wireless networks can be used in critical systems such as control loops, however, the unreliable nature of the wireless medium, coupled with their large scale and "ad-hoc" nature raise some of the most important challenges for low-power wireless research over the next 5-10 years.

Through the involvement of team members in standardization activities, the protocols and techniques will be proposed for the standardization process with a view to becoming the *de-facto* standard for wireless industrial process automation. Besides producing top level research publications and standardization activities, EVA intends this activity to foster further collaborations with industrial partners.

## 4.2. Environmental Monitoring

Today, outdoor WSNs are used to monitor vast rural or semi-rural areas and may be used to detect fires. Another example is detecting fires in outdoor fuel depots, where the delivery of alarm messages to a monitoring station in an upper-bounded time is of prime importance. Other applications consist in monitoring the snow melting process in mountains, tracking the quality of water in cities, registering the height of water in pipes to foresee flooding, etc. These applications lead to a vast number of technical issues: deployment strategies to ensure suitable coverage and good network connectivity, energy efficiency, reliability and latency, etc.

We work on such applications in an associate team "REALMS" comprising members from EVA, the university of Berkeley and the university of Michigan.

## 4.3. The Internet of Things

The general agreement is that the Internet of Things (IoT) is composed of small, often battery-powered objects which measure and interact with the physical world, and encompasses smart home applications, wearables, smart city and smart plant applications.

It is absolutely essential to (1) clearly understand the limits and capabilities of the IoT, and (2) develop technologies which enable user expectation to be met.

The EVA team is dedicated to understanding and contributing to the IoT. In particular, the team will maintain a good understanding of the different technologies at play (Bluetooth, IEEE 802.15.4, WiFi, cellular), and their trade-offs. Through scientific publications and other contributions, EVA will help establishing which technology best fits which application.

[0] Highway Addressable Remote Transducer

## 4.4. Military, Energy and Aerospace

Through the HIPERCOM project, EVA has developed cutting-edge expertise in using wireless networks for military, energy and aerospace applications. Wireless networks are a key enabling technology in the application domains, as they allow physical processes to be instrumented (e.g. the structural health of an airplane) at a granularity not achievable by its wired counterpart. Using wireless technology in these domains does however raise many technical challenges, including end-to-end latency, energy-efficiency, reliability and Quality of Service (QoS). Mobility is often an additional constraint in energy and military applications. Achieving scalability is of paramount importance for tactical military networks, and, albeit to a lesser degree, for power plants. EVA will work in this domain.

Smart cities share the constraint of mobility (both pedestrian and vehicular) with tactical military networks. Vehicular Ad-hoc NETworks (VANETs) will play an important role in the development of smarter cities.

The coexistence of different networks operating in the same radio spectrum can cause interference that should be avoided. Cognitive radio provides secondary users with the frequency channels that are temporarily unused (or unassigned) by primary users. Such opportunistic behavior can also be applied to urban wireless sensor networks. Smart cities raise the problem of transmitting, gathering, processing and storing big data. Another issue is to provide the right information at the place where it is most needed.

## 4.5. Emergency Applications

In an "emergency" application, heterogeneous nodes of a wireless network cooperate to recover from a disruptive event in a timely fashion, thereby possibly saving human lives. These wireless networks can be rapidly deployed and are useful to assess damage and take initial decisions. Their primary goal is to maintain connectivity with the humans or mobile robots (possibly in a hostile environment) in charge of network deployment. The deployment should ensure the coverage of particular points or areas of interest. The wireless network has to cope with pedestrian mobility and robot/vehicle mobility. The environment, initially unknown, is progressively discovered and may contain numerous obstacles that should be avoided. The nodes of the wireless network are usually battery-powered. Since they are placed by a robot or a human, their weight is very limited. The protocols supported by these nodes should be energy-efficient to maximize network lifetime. In such a challenging environment, sensor nodes should be replaced before their batteries are depleted. It is therefore important to be able to accurately determine the battery lifetime of these nodes, enabling predictive maintenance.

## 4.6. Types of Wireless Networks

The EVA team will distinguish between opportunistic communication (which takes advantage of a favorable state) and collaborative communication (several entities collaborate to reach a common objective). Furthermore, determinism can be required to schedule medium access and node activity, and to predict energy consumption.

In the EVA project, we will propose **self-adaptive wireless networks** whose evolution is based on:

- optimization to minimize a single or multiple objective functions under some constraints (e.g. interference, or energy consumption in the routing process).
- machine learning to be able to predict a future state based on past states (e.g. link quality in a wireless sensor network) and to identify tendencies.

The types of wireless networks encountered in the application domains can be classified in the following categories.

### *4.6.1. Wireless Sensor and Mesh Networks*

Standardization activities at the IETF have defined an "upper stack" allowing low-power mesh networks to be seamlessly integrated in the Internet (6LoWPAN), form multi-hop topologies (RPL), and interact with other devices like regular web servers (CoAP).

Major research challenges in sensor networks are mostly related to (predictable) power conservation and efficient multi-hop routing. Applications such as monitoring of mobile targets, and the generalization of smart phone devices and wearables, have introduced the need for WSN communication protocols to cope with node mobility and intermittent connectivity.

Extending WSN technology to new application spaces (e.g. security, sports, hostile environments) could also assist communication by seamless exchanges of information between individuals, between individuals and machines, or between machines, leading to the Internet of Things.

### *4.6.2. Deterministic Low-Power Networks*

*Wired* sensor networks have been used for decades to automate production processes in industrial applications, through standards such as HART. Because of the unreliable nature of the wireless medium, a wireless version of such industrial networks was long considered infeasible.

In 2012, the publication of the IEEE 802.15.4e standard triggered a revolutionary trend in low-power mesh networking: merging the performance of industrial networks, with the ease-of-integration of IP-enabled networks. This integration process is spearheaded by the IETF 6TiSCH working group, created in 2013. A 6TiSCH network implements the IEEE 802.15.4e TSCH protocol, as well as IETF standards such as 6LoWPAN, RPL and CoAP. A 6TiSCH network is synchronized, and a communication schedule orchestrates all communication in the network. Deployments of pre-6TiSCH networks have shown that they can achieve over 99.999% end-to-end reliability, and a decade of battery lifetime.

The communication schedule of a 6TiSCH network can be built and maintained using a centralized, distributed, or hybrid scheduling approach. While the mechanisms for managing that schedule are being standardized by the IETF, which scheduling approach to use, and the associated limits in terms of reliability, throughput and power consumption remains entirely open research questions. Contributing to answering these questions is an important research direction for the EVA team.

### *4.6.3. MANETs and VANETs*

In contrast to routing, other domains in MANETs such as medium access, multi-carrier transmission, quality of service, and quality of experience have received less attention. The establishment of research contracts for EVA in the field of MANETs is expected to remain substantial. MANETs will remain a key application domain for EVA with users such as the military, firefighters, emergency services and NGOs.

Vehicular Ad hoc Networks (VANETs) are arguably one of the most promising applications for MANETs. These networks primarily aim at improving road safety. Radio spectrum has been ring-fenced for VANETs worldwide, especially for safety applications. International standardization bodies are working on building efficient standards to govern vehicle-to-vehicle or vehicle-to-infrastructure communication.

### *4.6.4. Cellular and Device-to-Device Networks*

We propose to initially focus this activity on spectrum sensing. For efficient spectrum sensing, the first step is to discover the links (sub-carriers) on which nodes may initiate communications. In Device-to-Device (D2D) networks, one difficulty is scalability.

For link sensing, we will study and design new random access schemes for D2D networks, starting from active signaling. This will assume the availability of a control channel devoted to D2D neighbor discovery. It is therefore naturally coupled with cognitive radio algorithms (allocating such resources): coordination of link discovery through eNode-B information exchanges can yield further spectrum usage optimization.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

- Finalist, best paper award at the Global IoT Summit 2018, for paper "Why Channel Hopping Makes Sense, even with IEEE 802.15.4 OFDM at 2.4 GHz".
- **Thomas Watteyne** identified as "Key Innovator" by the European Commission's Innovation Radar, category "commitment" for the innovation "Online platform of testing tools for the Internet of Things".

### *5.1.2. Transfer*

- Creationg of the Wattson Element startup, which commercializes the Falco solution (https://wefalco.fr/).
- Publication of RFC8480

# 6. New Software and Platforms

## 6.1. OpenWSN

KEYWORDS: Internet of things - 6TiSCH - 6LoWPAN - CoAP
FUNCTIONAL DESCRIPTION: OpenWSN is an open-source implementation of a fully standards-based protocol stack for the Internet of Things. It has become the de-facto implementation of the IEEE802.15.4e TSCH standard, has a vibrant community of academic and industrial users, and is the reference implementation of the work we do in the IETF 6TiSCH standardization working group.

- Partner: University of California Berkeley
- Contact: Thomas Watteyne
- URL: http://www.openwsn.org/

## 6.2. 6TiSCH Simulator

*High-level simulator of a 6TiSCH network*
KEYWORDS: Network simulator - 6TiSCH
FUNCTIONAL DESCRIPTION: The simulator is written in Python. While it doesn't provide a cycle-accurate emulation, it does implement the functional behavior of a node running the full 6TiSCH protocol stack. This includes RPL, 6LoWPAN, CoAP and 6P. The implementation work tracks the progress of the standardization process at the IETF.

- Contact: Malisa Vucinic

## 6.3. Argus

KEYWORDS: Cloud - Low-Power WIreless - Sniffer
FUNCTIONAL DESCRIPTION: There are three piece to the Argus:

The Argus Probe is the program which attaches to your low-power wireless sniffer and forwards its traffic to the Argus Broker.

The Argus Broker sits somewhere in the cloud. Based on MQTT, it connect Argus Probes with Argus Clients based on a pub-sub architecture.

Several Argus Clients can the started at the same time. It is a program which subscribes to the Argus Broker and displays the frames in Wireshark.

- Contact: Rémy Leone

## 6.4. SolSystem

*Sensor Object Library System*
KEYWORDS: Low-Power WIreless - Back-End System - SmartMesh IP
FUNCTIONAL DESCRIPTION: The source code is composed of the definition of the SOL structure (https://github.com/realms-team/sol), the code that runs on the manager (https://github.com/realms-team/solmanager, written in Python) and the code that runs on the server receiving the data (https://github.com/realms-team/solserver, written in Python)

- Contact: Keoma Brun-Laguna
- URL: http://www.solsystem.io/

## 6.5. 6TiSCH Wireshark Dissector

KEYWORDS: 6TiSCH - Wireshark
FUNCTIONAL DESCRIPTION: Implementation on the dissectors is done through an open-source repository, stable code is regularly contributed back to the main Wireshark code base.

- Contact: Jonathan Munoz

## 6.6. F-Interop

*Remote Conformance and Interoperability Tests for the Internet of Thing*
KEYWORDS: Interoperability - Iot - Conformance testing - Standardization

- Partners: UPMC - IMEC - ETSI - EANTC - Mandat International - Digital Catapult - University of Luxembourg - Device Gateway
- Contact: Rémy Leone

## 6.7. Mercator

KEYWORDS: Deployment - Low-Power WIreless - Testbeds - Connectivity
FUNCTIONAL DESCRIPTION: The firmware is written as part of the OpenWSN project. Scripts and analysis tools are written in Python.

- Contact: Keoma Brun-Laguna

# 7. New Results

## 7.1. From SmartMarina to Falco

**Participants:** Keoma Brun-Laguna, Thomas Watteyne.

SmartMarina project (http://smartmarina.org/) was a technical project in 2017 to study the feasibility of using the wireless technology developed at Inria-EVA for marina management. In 2018, the Wattson Elements company was born, which now commercializes the Falco solution (https://wefalco.fr/).


*Figure 1. Screenshot of the Falco promotional video, https://youtu.be/35HdoFLrCf0.*

## 7.2. 6TiSCH Standardization

**Participants:** Malisa Vucinic, Jonathan Muñoz, Tengfei Chang, Yasuyuki Tanaka, Thomas Watteyne.

The standardization work at 6TiSCH remains a strong federator of the work done in the team. In 2018, the working group published the specification of the 6TiSCH Operation Sublayer (6top) Protocol, RFC8480. Work is also ongoing in the fragment forwarding space, where we are working on how ot efficient forward long IPv6 packets which are fragmented to fit in short IEEE 802.15.4 frames.

## 7.3. 6TiSCH Security

**Participants:** Malisa Vucinic, Thomas Watteyne.

The security work of Inria-EVA revolves around 6TiSCH networks and is a continuation of the efforts started during the H2020 ARMOUR project. The work focused on stabilizing the "Minimal Security" solution that has now passed the working group last call in the IETF and is pending finals reviews before being published as an RFC. The solution that is standardized enables secure network access and configuration of 6TiSCH devices under the assumption that they have been provisioned with a secret key. Ongoing work extends this solution to support true zero-configuration network setup, under the assumption that the devices have been provisioned with certificates at manufacturing time.

## 7.4. 6TiSCH Benchmarking

**Participants:** Malisa Vucinic, Tengfei Chang, Yasuyuki Tanaka, Thomas Watteyne.

With the pure 6TiSCH standardizes coming to an end, the focus of the group is moving towards benchmarking how well it works. This has results in the following action. Although seemingly different, they all contribute to the overall goal of better understand (the performance of) 6TiSCH.

We have built and put online the OpenTestbed, a collection of 80 OpenMote B boards deployed in 20 “pods”. These allow us to test the performance of the OpenWSN firmware in a realistic setting. The testbed is depicted in Fig. 2. You can access its management interface at http://testbed.openwsn.org/.


*Figure 2. The OpenTestbed deployed in Inria Paris since July 2018.*

A tool complementary to the testbed is the 6TiSCH simulator (https://bitbucket.org/6tisch/simulator) which Yatsuyiki Tanaka is leading. The simulator now represents exactly the behavior of the 6TiSCH protocol stack, and has been a catalyst for benchmarking activities around 6TiSCH.

Beyond Inria, the benchmarking activity around 6TiSCH is a hot topic, with projects such as the 6TiSCH Open Data Action (SODA, http://www.soda.ucg.ac.me/), the IoT Benchmarks Initiative (https://www.iotbench.ethz.ch/), and the Computer and Networking Experimental Research using Testbeds (CNERT) workshop at INFOCOM, all of which Inria-EVA is very involved in.

## 7.5. IoT and Wireless Sensor Networks

More than 50 billions of devices will be connected in 2020. This huge infrastructure of devices, which is managed by highly developed technologies, is called Internet of Things (IoT). The latter provides advanced services, and brings economical and societal benefits. This is the reason why engineers and researchers of both industry and scientific communities are interested in this area. The Internet of Things enables the interconnection of smart physical and virtual objects, managed by highly developed technologies. WSN

(Wireless Sensor Network), is an essential part of this paradigm. The WSN uses smart, autonomous and usually limited capacity devices in order to sense and monitor their environment.

### *7.5.1. Distributed Scheduling for IEEE 802.15.4e TSCH networks*

**Participants:** Yasuyuki Tanaka, Pascale Minet, Thomas Watteyne.

Since the scheduling algorithm is not standardized for IEEE 802.15.4e TSCH networks, many scheduling algorithms have been proposed. Most of them are centralized, few are distributed. Among the distributed scheduling algorithms, many rely on assumptions that may be violated by real deployments. This violation usually leads to conflicting transmissions of application data, decreasing the reliability and increasing the latency of data delivery. Others require a processing complexity that cannot be provided by sensor nodes of limited capabilities. Still others are unable to adapt quickly to traffic or topology changes, or are valid only for small traffic loads.

In the study funded by the Inria ADT DASMU (Action de Developement Technologique Distributed Adaptive Scheduling for MUltichannel wireless sensor networks), we focus on a distributed scheduling algorithm that relies on realistic assumptions, does not require complex computation, is valid for any traffic load, is adaptive and compliant with the standardized protocols used in the 6TiSCH working group at IETF.

First results have been obtained and an intensive simulation campaign made with the 6TiSCH simulator has provided comparative performance results. Our proposal outperforms MSF, the 6TiSCH Minimal Scheduling Function, in terms of end-to-end latency and end-to-end packet delivery ratio. More evaluations are needed to improve the proposal (e.g. less packet drops during transient situations, less overhead) in terms of scheduled cells).

### *7.5.2. IoT and IEEE 802.15.4e TSCH networks*

**Participants:** Pascale Minet, Ines Khoufi, Zied Soua.

In 2018, we focus on how an IEEE 802.15.4e is autonomously built and how nodes join the network.

To join the TSCH network, a device randomly selects a physical channel used by this network and listens to a beacon advertising this network. Since the physical channel on which the beacon is broadcast changes at each beacon slot due to channel hopping, the joining device will eventually hear a beacon sent by one of its neighbors. Upon receipt of a valid beacon, this device gets synchronized with the TSCH network.

In this study, we focus on the time needed by a node to detect a beacon sent by a TSCH network, as well as on the time needed to build a TSCH network. These times are important for industrial applications where new nodes are inserted progressively, or when failed nodes are replaced. Both times highly depend on the beacon advertisement policy, policy that is not specified in the standard and is under the responsibility of a layer upper than the MAC one. Since beacons are broadcast, they are lost in case of collisions: the vital information they carry is lost. The main problem is how to avoid collisions between two devices that are not neighbors.

That is why we propose the Enhanced Deterministic Beacon Advertising algorithm, called EDBA, that ensures a collision-free advertising of beacons. Since the beacon cells are fairly distributed in the slotframe, the average joining time is minimized. The behavior of a joining node has been modeled by a Markov chain from which the average joining time is computed, taking into account the reliability of wireless links. An intensive performance evaluation based on NS3 simulations allows us to validate this model and conclude on the very good performance of EDBA, even when compared with MBS, considered as the best advertising algorithm in the literature. These results have been published in the Annals of Telecommunications, [10].

### *7.5.3. UAV-based Data Gathering*

**Participants:** Nadjib Achir ( Paris 13), Tounsia Djamah, Paul Muhlethaler, Celia Tazibt ( Paris 13).

The recent advances in wireless sensors and Unmanned Aerial Vehicles have created new opportunities for environmental control and low cost aerial data gathering. We propose to use an Unmanned Aerial Vehicle (UAV) for data gathering [36]. Basically, we have proposed a method for UAV path planning based on virtual forces and potential fields. In addition, and more importantly, we present a new approach to compute the attractive forces of the potential field.

We use as our starting point the idea used by Pereira of using a potential field approach. However, we extend this work by considering that each cell in the area apply an attractive force on the drone, not only the deployed sensors. We compared our results with those obtained with Pereira's method and we obtained better performance in terms of data collection time. In other words, for the same period of time our method collect more data. The second advantage of our approach is that it leads to a significant reduction in the distance that the drone must travel.

### 7.5.4. Towards evaluating Named Data Networking for the IoT: A framework for OMNeT++

**Participants:** Amar Abane, Samia Bouzefrane ( Cnam), Paul Muhlethaler.

Named Data Networking is a promising architecture for emerging Internet applications such as the Internet of Things (IoT). Many studies have already investigated how NDN can be an alternative for IP in future IoT deployments. However, NDN-IoT propositions need accurate evaluation at network level and system level as well. We introduce an NDN framework for OMNeT++ [29]. Designed for low-end devices and gateways of the IoT, the framework is capable of simulating NDN scenarios at the boundary of the network and the system. The framework implementation is presented and used to study a typical aspect of NDN integration in IoT devices.

### 7.5.5. Evaluation of LORA with stochastic geometry

**Participants:** Bartek Blaszczyszyn ( Dyogen), Paul Muhlethaler.

We present a simple, stochastic-geometric model of a wireless access network exploiting the LoRA (Long Range) protocol, which is a non-expensive technology allowing for long-range, single-hop connectivity for the Internet of Things. We assume a space-time Poisson model of packets transmitted by LoRA nodes to a fixed base station. Following previous studies of the impact of interference, we assume that a given packet is successfully received when no interfering packet arrives with similar power before the given packet payload phase, see [39]. This is as a consequence of LoRa using different transmission rates for different link budgets (transmissions with smaller received powers use larger spreading factors) and LoRa intra-technology interference treatment. Using our model, we study the scaling of the packet reception probabilities per link budget as a function of the spatial density of nodes and their rate of transmissions. We consider both the parameter values recommended by the LoRa provider, as well as proposing LoRa tuning to improve the equality of performance for all link budgets. We also consider spatially non-homogeneous distributions of LoRa nodes. We show also how a fair comparison to non-slotted Aloha can be made within the same framework.

### 7.5.6. Position Certainty Propagation: A location service for MANETs

**Participants:** Abdallah Sobehy, Paul Muhlethaler, Eric Renault ( Telecom Sud-Paris).

Localization in Mobile Ad-hoc Networks (MANETs) and Wireless Sensor Networks (WSNs) is an issue of great interest, especially in applications such as the IoT and VANETs. We propose a solution that overcomes two limiting characteristics of these types of networks. The first is the high cost of nodes with a location sensor (such as GPS) which we will refer to as anchor nodes. The second is the low computational capability of nodes in the network. The proposed algorithm [28] addresses two issues; self-localization where each non-anchor node should discover its own position, and global localization where a node establishes knowledge of the position of all the nodes in the network. We address the problem as a graph where vertices are nodes in the network and edges indicate connectivity between nodes. The weights of edges represent the Euclidean distance between the nodes. Given a graph with at least three anchor nodes and knowing the maximum communication range for each node, we are able to localize nodes using fairly simple computations in a moderately dense graph.

## 7.6. Industry 4.0 and Low-Power Wireless Meshed Networks

### *7.6.1. Deterministic Networking for the Industrial Internet of Things (IIoT)*

**Participants:** Keoma Brun-Laguna, Thomas Watteyne, Pascale Minet.

The Internet of Things (IoT) connects tiny electronic devices able to measure a physical value (temperature, humidity, etc.) and/or to actuate on the physical world (pump, valve, etc). Due to their cost and ease of deployment, battery-powered wireless IoT networks are rapidly being adopted.

The promise of wireless communication is to offer wire-like connectivity. Major improvements have been made in that sense, but many challenges remain as industrial application have strong operational requirements. This section of the IoT application is called Industrial IoT (IIoT).

The main IIoT requirement is reliability. Every bit of information that is transmitted in the network must not be lost. Current off-the-shelf solutions offer over 99.999% reliability.

Then come latency and energy-efficiency requirements. As devices are battery-powered, they need to consume as little as possible to be able to operate during years. The next step for the IoT is to target time-critical applications.

Industrial IoT technologies are now adopted by companies over the world, and are now a proven solution. Yet, challenges remain and some of the limits of the technologies are still not fully understood. In his PhD Thesis, Keoma Brun-Laguna addresses TSCH-based Wireless Sensor Networks and studies their latency and lifetime limits under real-world conditions.

We gathered 3M network statistics 32M sensor measurements on 11 datasets with a total of 170,037 mote hours in real-world and testbeds deployments. We assembled what we believed to be the largest dataset available to the networking community.

Based on those datasets and on insights we learned from deploying networks in real-world conditions, we study the limits and trade-offs of TSCH-based Wireless Sensor Networks. We provide methods and tools to estimate the network performances of such networks in various scenarios. We highlight the trade-off between short latency and long network lifetime. We believe we assembled the right tools for protocol designer to build deterministic networking to the Industrial IoT.

### *7.6.2. Industry 4.0 and IEEE 802.15.4e TSCH networks*

**Participants:** Pascale Minet, Ines Khoufi, Zied Soua.

By the year 2020, it is expected that the number of connected objects will exceed several billions devices. These objects will be present in everyday life for a smarter home and city as well as in future smart factories that will revolutionize the industry organization. This is actually the expected fourth industrial revolution, more known as Industry 4.0. In which, the Internet of Things (IoT) is considered as a key enabler for this major transformation. IoT will allow more intelligent monitoring and self-organizing capabilities than traditional factories. As a consequence, the production process will be more efficient and flexible with products of higher quality.

To produce better quality products and improve monitoring in Industry 4.0, strong requirements in terms of latency, robustness and power autonomy have to be met by the networks supporting the Industry 4.0 applications. The wireless TSCH (Time Slotted Channel Hopping) network specified in the e amendment of the IEEE 802.15.4 standard has many appealing properties. Its schedule of multichannel slotted data transmissions ensures the absence of collisions. Because there is no retransmission due to collisions, communication is faster. Since the devices save energy each time they do not take part in a transmission, the power autonomy of nodes is prolonged. Furthermore, channel hopping enables to mitigate multipath fading and interferences.

To increase the flexibility and the self-organizing capacities required by Industry 4.0, the networks have to be able to adapt to changes. These changes may concern the application itself, the network topology by adding or removing devices, the traffic generated by increasing or decreasing the device sampling frequency, for instance. That is why the flexibility of the schedule ruling all network communications is needed.

In 2018, we show how a TSCH network can adapt to such changes. More precisely, we propose a solution ranging from network construction to data gathering. We show how a TSCH network is autonomously built, supports data gathering and is able to adapt to changes in network topology, traffic and application requirements.

The solution proposed preserves the merits of TSCH network, that can be listed hereafter. The time-slotted multichannel medium access enables parallel transmissions on several channels, leading to shorter latency and higher throughputs. In addition, channel hopping mitigates interference and multipath effects. Furthermore, since transmissions are scheduled, a conflict-free schedule is computed by the network coordinator (i.e. the CPAN). Hence, no collision occurs during data gathering. The absence of collision leads to a higher throughput, because there is no retransmission due to collisions. It also preserves nodes power autonomy.

This simple solution is based on the coexistence of several periodic slotframes. We distinguish three slotframes, which are the Beacon Slotframe, the Data Slotframe and the Shared Slotframe. The network schedule corresponds to the superposition of the three schedules given by each slotframe, where the slotframe with the highest priority wins.

This solution ensures a collision-free dissemination over the whole network. Beacons are broadcast in sequence by increasing depth of devices. This broadcast is also used to disseminate Data Schedules (new schedule or update).

In addition, this solution is adaptive. Topology, traffic or application changes are notified to the CPAN. Depending on the changes notified, the CPAN updates the current schedule or recomputes a new one. Shared slots are used to cope with unexpected events.

We compute the theoretical bounds with regard to key performance indicators and compare them with the values obtained by NS3 simulation. Simulation results confirm the theoretical upper bounds computed for network construction and data gathering. Hence, TSCH networks are able to adapt to traffic or topology changes in a reasonable time which is a strong requirement of Industry 4.0 applications. These results have been presented at the PEMWN 2018 conference in [26]. In some further work, we will study how to improve this delay to support the most demanding applications.

## 7.7. Machine Learning for an efficient and dynamic management of data centers

### *7.7.1. Data Analysis in Data Centers*

**Participants:** Eric Renault ( Telecom Sud-Paris), Selma Boumerdassi ( Cnam), Pascale Minet, Ines Khoufi.

In High Performance Computing (HPC), it is assumed that all machines are homogeneous in terms of CPU and memory capacities, and that the tasks making up the jobs have similar resource requests. It has been shown that this homogeneity relating both to machine capacity and workload, although generally valid for HPC, does no longer apply to data centers. This explains why the publication of data gathered in an operational Google data center over 29 days has aroused great interest among researchers.

It is crucial to have real traces of a Google data center publicly available that are representative of the functioning of real data centers. Our goal is to analyze the data collected and to draw useful conclusions about machines, jobs and tasks as well as resource usage. Our main results have been published in [25], [24] and can be summarized as follows:

- Although $92\%$ of machines have a CPU capacity of $0.5$, there are 10 machine configurations in the data center, each configuration is characterized by a pair $(CPU\ \ capacity, memory\ \ capacity)$. The most frequent configuration is supported by only $53\%$ of machines.
- Over the 29 days, all the machines in the data center that were removed, were restarted later after an off-period. $50\%$ of these periods have a duration less than or equal to 1000 seconds (i.e. $16.66$ minutes), suggesting a maintenance operation.

- The distribution of jobs per category reveals only one job, representing $0.002\%$, for the Infrastructure, $0.13\%$ of jobs for Monitoring, $9.91\%$ of jobs for Production, $56.30\%$ of jobs for Other, and $33.63\%$ of jobs for Free. $92.05\%$ of jobs have a single task. $95.75\%$ have fewer than 10 tasks. But 12 jobs have 5000 tasks and 114 jobs have around 1000 tasks.
- With regard to resource requests, $0.11\%$ of jobs have a memory request and a CPU request higher than or equal to $10\%$.
- $94.25\%$ of jobs wait less than 10 seconds before being scheduled. However, some of them wait for more than 1000 seconds. Such large values could be explained by the existence of placement constraints for the jobs, making them harder to place and schedule. $49\%$ of jobs have an execution time less than 100 seconds.

Such results are needed to validate or invalidate some simplifying assumptions that are usually made when reasoning about models, and make the models more accurate for jobs and tasks as well as for available machines. Having validated these models on real data centers, they can then be used for extensive evaluation of placement and scheduling algorithms and more generally for resource allocation (i.e. CPU and memory). These algorithms can then be applied in real data centers.

Another possible use of this data set is to consider it as a learning set in order to predict some feature of the data center, such as the workload of hosts or the next arrival of jobs.

### *7.7.2. Machine Learning for an Energy-Efficient Management of Data Centers*

**Participants:** Ruben Milocco ( University Of Camahue, Argentina), Pascale Minet, Eric Renault ( Telecom Sud-Paris), Selma Boumerdassi ( Cnam).

To limit global warming, all industrial sectors must make effort to reduce their carbon footprint. Information and Communication Technologies (ICTs) alone generate 2% of global CO2 emissions every year. Due to the rapid growth in Internet services, data centers have the largest carbon footprint of all ICTs. According to ARCEP (the French telecommunications regulator), Internet data traffic multiplied by 4.5 between 2011 and 2016. In order to support such a growth and maintain this traffic, data centers'energy consumption needs to be optimized. The problem of managing Data Centers (DC) and clouds optimally, in the sense that the demand is met with a minimal energy cost, remains a major issue. In this research, we evaluate the maximum energy saving that can be obtained in DCs by means of a proactive management of resources. The proposed management is based on models that predict resource requests.

Diverse approaches to obtain predictive models of DCs have been studied recently. Among the most popular methods with the comparatively lowest prediction errors are the predictive models of the ARMAX family. Hence, we study the predictive model given by the ARMAX family. We compare its performance with that of the Last Value (LV) model which predicts that the next value will be equal to the current one. To the best of our knowledge, there are no studies relating to the performance bounds that can be achieved using these models. In this research, we study the limits of the improvement in terms of energy cost that can be obtained using proactive strategies for DC management based on predictive models.

Using the Google dataset collected over a period of 29 days and made publicly available, we evaluate the largest benefit that can be obtained with those two predictors.

## 7.8. Protocols and Models for Wireless Networks - Application to VANETs

### *7.8.1. Predicting Vehicles Positions using Roadside Units: a Machine-Learning Approach*

**Participants:** Samia Bouzefrane ( Cnam), Soumya Banerjee ( Birla Institute Of Technology, Mesra), Paul Mühlethaler, Mamoudou Sangare.

We study positioning systems using Vehicular Ad Hoc Networks (VANETs) to predict the position of vehicles [35]. We use the reception power of the packets received by the Road Side Units (RSUs) and sent by the vehicles on the roads. In fact, the reception power is strongly influenced by the distance between a vehicle and a RSU. To predict the position of vehicles in this context, we adopt the machine learning methodology.

As a pre-requisite, the vehicles know their positions and the vehicles send their positions in the packets. The positioning system can thus perform a training sequence and build a model. The system is then able to handle a prediction request. In this request, a vehicle without external positioning will request its position from the neighboring RSUs. The RSUs which receive this request message from the vehicle will know the power at which the message was received and will study the positioning request using the training set. In this study, we use and compare three widely recognized techniques : K Nearest Neighbors (KNN), Support Vector Machine (SVM) and Random Forest. We study these techniques in various configurations and discuss their respective advantages and drawbacks. Our results show that these three techniques provide very good results in terms of position predictions when the error on the transmission power is small.

### *7.8.2. Predicting transmission success with Machine-Learning and Support Vector Machine in VANETs*

**Participants:** Samia Bouzefrane ( Cnam), Soumya Banerjee ( Birla Institute Of Technology, Mesra), Paul Mühlethaler, Mamoudou Sangare.

We study the use of the Support Vector Machine technique to estimate the probability of the reception of a given transmission in a Vehicular Ad hoc NETwork (VANET). The transmission takes place between a vehicle and a RoadSide Unit (RSU) at a given distance and with a given transmission rate. The RSU computes the statistics of the receptions and is able to compute the percentage of successful transmissions versus the distance between the vehicle and the RSU and the transmission rate. Starting from this statistic, a Support Vector Machine (SVM) scheme can produce a model. Then, given a transmission rate and a distance between the vehicle and the RSU, the SVM technique can estimate the probability of a succcessful reception. This probability can be used to build an adaptive technique which optimizes the expected throughput between the vehicle and the RSU. Instead of using transmission values of a real experiment, we use the results of an analytical model of CSMA that is customized for 1D VANETs. The model we adopt to perform this task uses a Matern selection process to mimic the transmission in a CSMA IEEE 802.11p VANET. With this model we obtain a closed formula for the probability of successful transmissions. Thus with these results we can train an SVM model and predict other values for other couples : distance, transmission rate. The numerical results we obtain show that SVM seems very suitable to predict the reception probability in a VANET.

### *7.8.3. TDMA scheduling strategies for vehicular ad hoc networks: from a distributed to a centralized approach*

**Participants:** Mohammed Hadded, Anis Laouiti ( Telecom Sud-Paris, Paul Mühlethaler.

We focus on vehicular safety applications based on the Dedicated Short Range Communication (DSRC) standard. We propose a new mechanism to alleviate channel congestion by reducing the beacons load while maintaining an accurate awareness level. Our scheme is based on the collective perception concept which consists in sharing perceived status information collected by vehicles equipped with different types of sensors (radars, lidars, cameras, etc.). To achieve our goal, we propose two main schemes [30]. The first one consists in implementing the collective perception capability on vehicles and adding a new category of status messages to share locally collected sensor data in order to reduce channels load and enhance vehicles' awareness. The second scheme concerns the accuracy level of the received information from the collective perception enabled vehicles by fixing a prior error threshold on the position. The method proposed is validated by simulations and the results obtained are compared to those of an application based on the traditional beaconing scheme of the IEEE802.11p standard. The simulations show that the proposed scheme is able to significantly reduce the load on the control channel incurred by the beacons and the packet error ratio for different network densities and built-in sensors characteristics.

### *7.8.4. A Collaborative Environment Perception Approach for Vehicular Ad hoc Networks*

**Participants:** Sadia Ingrachen, Nadjib Achir ( Paris 13), Paul Mühlethaler, Tounsia Djamah ( Paris 13), Amine Berqia ( Paris 13).

We focus on vehicular safety applications based on the Dedicated Short Range Communication (DSRC) standard. We propose a new mechanism to alleviate channel congestion by reducing the beacons load while

maintaining an accurate awareness level. Our scheme is based on the collective perception concept which consists in sharing perceived status information collected by vehicles equipped with different types of sensors (radars, lidars, cameras, etc.). To achieve our goal, we propose two main schemes [31]. The first one consists in implementing the collective perception capability on vehicles and adding a new category of status messages to share locally collected sensor data in order to reduce channels load and enhance vehicles' awareness. The second scheme concerns the accuracy level of the received information from the collective perception enabled vehicles by fixing a prior error threshold on the position. The method proposed is validated by simulations and the results obtained are compared to those of an application based on the traditional beaconing scheme of the IEEE802.11p standard. The simulations show that the proposed scheme is able to significantly reduce the load on the control channel incurred by the beacons and the packet error ratio for different network densities and built-in sensors characteristics.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

**Participants:** Pascale Minet, Ines Khoufi, Zied Soua.

In the framework of the CNES Launchers Research and Technology program, Inria and CNEs co-funded a study dealing with wireless sensor networks in a spatial environment. More precisely, this study deals with the improvement and performance evaluation of a solution of wireless sensor networksbased on the IEEE 802.15.4e standard of TSCH (Time Slotted Channel Hopping), operating in a spatial environment.

In space launch vehicles, a NASA study shows that the mass per channel of 0.45 kg for a wiring approach can be reduced to 0.09 kg for a wireless approach.8 A question arises: which wireless technology is able to meet the requirements of space launch vehicles in terms of latency, throughput and robustness. The IEEE 802.15.4e amendment has been designed to meet such requirements. More specifically, the Time Slotted Channel Hopping (TSCH) mode of the IEEE 802.15.4e standard that has been designed for industrial automation, process control and equipment monitoring, appears very promising for space launch vehicles. More precisely, the study for CNES deals with:

- Building an IEEE 802.15.4e TSCH network: see [11] the Acta Astronautica 2018 publication.
- Scheduling transmissions in an IEEE 802.15.4e TSCH network.
- Adapting the schedule to traffic or topology changes.

This study ended in July 2018 with very satisfying results.

## 8.2. Bilateral Grants with Industry

**Participants:** Thomas Watteyne, Felipe Moran.

Felipe Moran was awarded a 6-month EDF fellowship to conduct a 6-month internship around low-power wireless networking in extrement industrial environments. Details are confidential.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR*

- The GeoBot FUI project (https://geobot.fr/) is one of the most innovative, challenging and fun projects around wireless localization in the world today. It applies true innovation to a real-world problem, with a clear target application (and customer) in mind. The GeoBot partners are building a small robot (think of a matchbox-sized RC car) that will be inserted into a gas pipe, and move around it to map the location of the different underground pipes. Such mapping is necessary to prevent gas-related accidents, for example during construction. At the end of the project, this solution will be commercialized and used to map the network of gas pipe in France, before being used in worldwide. Each partner is in charge of a different aspect of the problem: robotics, analysis of the inertial data, visualization, etc. Inria is in charge of the wireless part. We will be equipping the robot with a wireless chip(set) in order to (1) communicate with the robot as it moves about in the pipes while standing on the surface, and (2) discover the relative location of the robot w.r.t. a person on the surface. Inria is evaluating different wireless technologies, benchmarking around ranging accuracy and capabilities to communicate. We start from off-the-shelf kits from different vendors and build a custom board, benchmark it, and integrate it with the other partners of the project.

### *9.1.2. Other collaborations*

- EVA has a collaboration with Orange Labs. **Thomas Watteyne** supervises the PhD of Mina Rady, which happens under a CIFRE agreement with Orange Labs.
- EVA has a collaboration with Vedecom. **Paul Muhlethaler** supervises Fouzi Boukhalfa's PhD funded by Vedecom. This PhD aims at studying low latency and high reliability vehicle-to-vehicle communication to improve roads safety.
- EVA has an ongoing collaboration with SODEAL company, which exploits the Cap d'Agde marina, as part of the SmartMarina project.

## 9.2. European Initiatives

### *9.2.1. FP7 & H2020 Projects*

The H2020 following projects are ongoing:

- H2020 SPARTA, Jan 2019 – December 2020.
- H2020 F-Interop, http://f-interop.eu/, Nov 2015 – Oct 2018.
- H2020 ARMOUR, https://www.armour-project.eu/, Feb 2016 – Jan 2018.

### *9.2.2. Collaborations with Major European Organizations*

Inria-EVA has collaboration in 2018 with ETSI (the European Telecommunications Standards Institute) to organize the F-Interop 6TiSCH 2 Interop Event on 2-4 February 2018 in Paris.

## 9.3. International Initiatives

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

*9.3.1.1. REALMS*

- Title: Real-Time Real-World Monitoring Systems
- International Partner (Institution - Laboratory - Researcher):
  - University of California Berkeley (United States) - Civil and Environmental Engineering - Steven Glaser
  - University of Michigan (United States) - Civil and Environmental Engineering - Branko Kerkez
- Start year: 2015
- See also: http://glaser.berkeley.edu et http://www-personal.umich.edu/~bkerkez/

- The Internet of Things revolution prompted the development of new products and standards; The IEEE 802.15.4e (2012) standard introduced the Time Synchronized Channel Hoping (TSCH) which can provide end-to-end reliability of 99.999 % and an energy autonomy of many years. This exceptional performance prompted the IETF to create the 6TISCH working group to standardize the integration of TSCH networks in the Internet. While the first experimental data have highlighted the great robustness of these networks, there is no data of a real network, accessible in real time, on a large scale and over a long period. Such data is needed to better model network performance and produce better products and standards. Teams of Professors Glaser and Kerkez are successfully deploying such networks to study mountain hydrology, monitor water quality and manage rainwater in urban environments. A model is missing to assist in the deployment and operation of these networks, as well as to monitor an operational network.

*9.3.1.2. DIVERSITY*

- Title: Measuring and Exploiting Diversity in Low-Power Wireless Networks
- International Partner (Institution - Laboratory - Researcher):
  - University of Southern California (United States) - Autonomous Networks Research Group (ANRG) - Bhaskar Krishnamachari
- Start year: 2016
- The goal of the DIVERSITY associate team is to develop the networking technology for tomorrow's Smart Factory. The two teams comes with a perfectly complementary background on standardization and experimentation (Inria-EVA) and scheduling techniques (USC-ANRG). The key topic addressed by the joint team will be networking solutions for the Industrial Internet of Things (IIoT), with a particular focus on reliability and determinism.

### *9.3.2. Inria International Partners*

*9.3.2.1. Declared Inria International Partners*

Inria-EVA has a long-standing Memorandum of Understanding with the OpenMote company (http://www.openmote.com/), which runs until 2020. OpenMote emerged as a spin-off of the OpenWSN project, co-lead by **Thomas Watteyne** and Prof. Xavier Vilajosana, Professor at the Open University of Catalonia and Chief Technical Officer at OpenMote.

The collaboration has been ongoing since 2012 and at the time of writing has resulted in:

- Joint academic publications, including 7 journal articles, 1 letter, 1 book chapter, 5 conference papers, 2 tutorials and invited talks.
- Joint standardization activities, in particular in the IETF 6TiSCH working group, co-chaired by **Thomas Watteyne** and for which Prof. Xavier Vilajosana is a key contributor. This activity has resulted in the joint participation in 12 IETF face-to-face meetings, joint participation in over 100 audioconferences, co-authorship of 3 Internet-Drafts and joint organization of 2 interop events.
- Joint software development, as both institutions closely collaborate in the maintenance, development, promotion and research along the OpenWSN project, including the development of the protocol stack, the integration of novel hardware technologies, the support to the community and the participation in standardization activities and interoperability events.

This MOU is NOT a commitment of funds by any part.

*9.3.2.2. Informal International Partners*

The Inria-EVA collaborates extensively with Prof. Pister's group at UC Berkeley on the OpenWSN and Smart Dust projects. This activity translated into several members of the Pister team visiting Inria-EVA and vice-versa in 2018.

*9.3.2.3. International Initiatives*

Inria-EVA participates in the IoT Benchmarks Initiative (https://www.iotbench.ethz.ch/)

Inria-EVA will be participating in 2019 in the WirelessWine SticAm-Sud project.

## 9.4. International Research Visitors

### *9.4.1. Visits of International Scientists*

1. **Prof. Xavi Vilajosana (UOC/OpenMote)** (26-30 November 2018) working on OpenMote B bring-up with Tengfei Chang and **Thomas Watteyne**
2. **Brian Gregory Kilberg (UC Berkeley)** (11-18 September 2018) working on OpenWSN/ROS integration with **Thomas Watteyne** and Tengfei Chang
3. **Prof. Xavi Vilajosana (UOC/OpenMote)** (24-28 June 2018) working on F-Interop 6TiSCH with **Thomas Watteyne** and Tengfei Chang
4. **Pablo Modernell (UOC)** (28 May – 1 June 2018) working on F-Interop with Tengfei Chang and **Thomas Watteyne**
5. **Malisa Vucinic (U Montenegro)** (9 -16 March 2018) working on 6TiSCH Security with **Thomas Watteyne**
6. **Lance Doherty (Analog Devices)** (8-9 February 2018) working on SmartMesh IP with **Thomas Watteyne**
7. **Malisa Vucinic (U Montenegro)** (29 January-16 February 2018) working on 6TiSCH Security with **Thomas Watteyne**

### *9.4.2. Internships*

1. **Felipe Moran**, MSc intern from ENSTA ParisTech (1 September 2017 – 31 August 2018), EDF fellow, Research Topic: mote feeding habits, SmartMesh IP, Advisor: Thomas Watteyne
2. **Fabian Rincon Vija**, MSc intern from ENSTA ParisTech (14 May – 31 August 2018), Research Topic: Extension of F-Interop to IEEE 802.15.4 sub-GHz, Advisor: Thomas Watteyne
3. **Marcelo Augusto Ferreira**, MSc intern from ENSTA ParisTech (1 May – 31 August 2018), Research Topic: Measuring Energy Consumption in F-Interop, Advisor: Thomas Watteyne
4. **Imene Ben Haddada**, Using Support Vector Machine for Positioning Services in Vehicle Ad-hoc NETworks (ENSI- Tunisia), March-July 2018.
5. **Khalifa Hadded**, Generation of positioning data in Vehicle Ad-hoc NETworks (ENSI- Tunisia), March-July 2018.
6. **Zied Soua**, Formation d'un réseau TSCH
   et ordonnancement de ses communications dans le cadre de l'IoT industriel, (INSAT- Tunisia), February-July 2018.

### *9.4.3. Visits to International Teams*

*9.4.3.1. Research Stays Abroad*

- **Thomas Watteyne** spent the month of August 2017 at UC Berkeley, working with Prof. Glaser on the SnowHow project, and with Prof. Pister on Smart Dust and OpenWSN.
- Tengfei Chang spent the month of July 2017 in California working with Prof. Pister working on Smart Dust UC Berkeley, and Prof. Krishnamachari working on testbed deployment at the Univesity of Southern California.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organization*

*10.1.1.1. General Chair, Scientific Chair*

- **Thomas Watteyne** is TPC co-chair of the 6th International Workshop on Computer and Networking Experimental Research using Testbeds (CNERT), held in conjunction with IEEEINFOCOM, Paris, France, 29 April-2 May, 2019. More information at https://infocom2019.ieee-infocom.org/cnert-computer-and-networking-experimental-research-using-testbeds.
- **Paul Muhlethaler** was general co-chair with Eric Renault of the first conference of application of Machine Learning for Networks (MLN 2018 27-28 November 2018), conference hosted by Inria Paris.
- **Pascale Minet** was general co-chair with Leila Saidane from ENSI (Tunisia) of the PEMWN 2018 conference, the 7th IFIP/IEEE international conference on Performance Evaluation and Modeling of Wired and Wireless Networks, technically co-sponsored by IFIP WG6.2 and IEEE ComSoc (see https://sites.google.com/site/pemwn2018/). This conference was held in Toulouse (IUT of Blagnac), the 26th, 27th and 28th of September 2018. The organization co-chairs were Thierry Val, Adrien Van Den Bossche, and Rejane Dalce. Three tutorials were given:
  - *D2D communications in 5G networks: opportunities and challenges* by Salah Eddine Elayoubi, Centrale Supelec, Paris.
  - *Drone aided networks: from collecting data to connecting people* by Riadh Dhaou, Institut National Polytechnique de Toulouse (INPT).
  - *Continuity of the Positioning Service* by François Spies, University Bourgogne Franche-Comte.
- **Thomas Watteyne** was co-chair of the 6TiSCH 2 Plugtests, Inria, Paris,2-4 February 2018.

*10.1.1.2. Member of the Organizing Committees*

- **Paul Muhlethaler** organized the DGA Inria workshop on Artificial Intelligence for telecommunications and networks in May 2018.

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Chair of Conference Program Committees*

- **Paul Muhlethaler** was in Steering committee member of MobileHealth Workshop 2018.
- **Anis Louiti** was in Steering committee member of MobileHealth Workshop 2018.

*10.1.2.2. Member of the Conference Program Committees*

- **Pascale Minet**
  - CoRes 2018, 3emes Rencontres Francophones sur la Conception de Protocoles,
  - DCNET 2018, 9th International Conference on Data Communication Networking, July 2018,
  - ETFA 2018, 22th IEEE International Conference on Emerging Technologies & Factory Automation, September 2018,
  - EUSPN 2018, 9th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN), November 2018,
  - GlobeCom 2018, IEEE Global Communications Conference, December 2018.
  - MLN 2018, Machine Learning for Networking, November 2018,
  - MSPN 2018, 4th International Conference on Mobile, Secure and Programmable Networking, June 2018,
  - PEMWN 2018, 7th International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks, September 2018,
  - PECCS 2018, 8th International conference on Pervasive and Embedded Computing and Communication Systems, July 2018,

  - SITIS 2018, 14th International Conference on Signal Image Technology & Internet Based Systems, November 2018,
  - VTC 2018, 87th IEEE Vehicular Technology Conference, June 2018,
  - Wireless Days 2018, IFIP/IEEE Wireless Days, March 2018,
  - WiSEE 2018, 6th IEEE International Conference on Wireless for Space and Extreme Environments, October 2018.
- **Paul Muhlethaler**:
  - ITST 2018 15- 17 October, Lisbon, Portugal ,
  - ISCC 2018, 25-28 June 2018, Natal Brazil,
  - Mownet 2018, International Conference on Selected Topics in Mobile & Wireless Networking, 20 -22 June 2018,
  - PEMWN 2018, 7th International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks, 26 - 28 November 2018, Toulouse,
  - Wireless Days, IFIP/IEEE Wireless Days 3 - 5 April 2018, Dubai,
  - WiOpt 2018, 7 - 11 March 2018 Shangai, China.

### *10.1.3. Journal*

*10.1.3.1. Reviewer - Reviewing Activities*

- **Paul Muhlethaler**
  - Reviewer Ad Hoc Networks Journal (Elsevier),
  - Reviewer Annals of telecommunications,
  - Reviewer International Journal of Distributed Sensor Networks. Hindawi,
  - Reviewer IEEE Transactions on Information Theory,
  - Reviewer IEEE Transactions on Vehicular Technology,
  - Reviewer IEEE Transactions on Wireless Communications.
- **Pascale Minet**
  - Acta Astronautica,
  - Ad Hoc Networks,
  - Annals of Telecommunications,
  - Computer Communications,
  - Computer Networks,
  - Engineering Applications of Artificial Intelligence,
  - IEEE Access,
  - IEEE Internet of Things,
  - IEEE Transactions on Mobile Computing,
  - IEEE Transactions on Green Communications and Networking,
  - IEEE Transactions on Industrial Electronics,
  - IEEE Transactions on Industrial Informatics,
  - International Journal of Communication Systems,
  - International Journal of Ad Hoc and Ubiquitous Computing,
  - Sensors Journal,
  - Wireless Networks.
- **Thomas Watteyne**

  - MDPI Sensors, 2018.
- Nadjib Achir
  - Reviewer Sensor Networks (MDPI)
  - Reviewer Wireless Communications and Mobile Computing (Wiley)
  - Reviewer Internet of Things Journal (IEEE)
  - Reviewer Ad Hoc Networks Journal (Elsevier)
- Selma Boumerdassi
  - Reviewer Ad Hoc Networks Journal (Elsevier);
  - Reviewer The journal of Future Generation Computer Systems (Elsevier).
- Samia Bouzefrane
  - The International Journal of Computer and Telecommunications Networking (Elsevier),
  - The IEEE Transactions on Mobile Computing,
  - The Information and Software Technology Journal (Elsevier)
  - The Springer Multimedia Tools and Application Journal
  - The ACM Transaction on Internet Technology
  - The Concurrency and Computation Practice and Experience Journal
  - the Journal of Systems and Software (Elsevier)

### *10.1.4. Invited Talks*

- **Thomas Watteyne**
  - Panelist on "New Trends for IoT and Satellite Networks", talking about "Industrial IoT, A Reality Check – Standards, Products and Research Challenges" at IoT Week 2018, Bilbao, Spain, 6 June 2018.
  - Boostez la Performance the vos Objects Connectes. Inria Tech talk, French Tech central, Station F, Paris, France. 24 October 2018.
  - Getting Your Hands Dirty with the Industrial IoT and SmartMesh IP. Summer School on Dependable IoT. TU Graz, Graz, Austria. 20 July 2018.
  - SmartMesh IP. Captronic/Arrow mesh networking day. ESSIE, Paris, France. 25 October 2018.
  - The Internet of (Important) Things. French Tech Central, Station F, Paris, France. 24 October 2018.
  - "IoT research, standardization and interop using testbeds", Fed4FIRE Engineering Conference, Brugge, Belgium, 8-9 October 2018.
  - From Research, to Product, to Standardization: A Journey into TSCH. TU Graz, Graz, Austria, 19 July 2018.
  - Reality Check on IoT Solutions producing data for people to analyze. Boston Consultancy Group semine series, Station F, Paris, France. 7 July 2018.
  - (keynote) Industrial IoT, A Reality Check : Standards, Products and Research Challenges. IIoT Workshop, Strasbourg, France, 3 July 2018.
  - "IPv6 over the TSCH mode of IEEE 802.15.4e: overview of standardization, tooling, open-source initiative and commercial products", Workshop on Design, Deployment and Testing of Internet of Things Technologies (DDT-IoT), IEEE BalkanCom, Podgorica, Montenegro, 8 June 2018.
  - "Industrial IoT, A Reality Check: Standards, Products and Research Challenges", IoT Week, Bilbao, Spain, 4-8 June 2018.

- A Turn-Key Solution for Real-World IoT, presented together with Keoma Brun-Laguna. LIRIMA Workshop on Smart Agriculture in Africa. 3-4 April 2018.
- Intro to Dr. Malisa Vucinic' keynote "The Devil is in the Detail: How a Real-World IoT Technology is Made – IETF 6TiSCH" at the Information Technology Conference, Montenegro, 20 February 2018.

### *10.1.5. Leadership within the Scientific Community*

**Thomas Watteyne** co-chairs the IETF 6TiSCH standardization group.

### *10.1.6. Scientific Expertise*

**Thomas Watteyne** regularly consults with major player in the (Industrial) IoT space.

### *10.1.7. Research Administration*

- **Thomas Watteyne** is member of the Inria-Paris "Commission de Developpement Technologique", since 2018, where we ensure Inria project teams get sufficient engineering resources to change the world.
- **Paul Muhlethaler** is member of the Inria-Paris "Comite de Centre", since 2016 (suppléant of Michel Kern).
- **Thomas Watteyne** is member of the Inria-Paris "Comite de Centre", since 2016, where we work on making sure Inria-Paris will always remain one of the greatest places to work at!

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

- Intensive 1-week course on IoT, with associated hands-on labs. ENSTA ParisTech. Graduate level. **Thomas Watteyne** and Ziran Zhang. 9-12 October 2017.
- 1/2-day crash course on the Industrial IoT. Telecom ParisTech. Graduate level. **Thomas Watteyne**. 28 September 2017..
- 6-week course on IoT, with associated hands-on labs. ENSTA ParisTech. Undergraduate level. **Thomas Watteyne**, Keoma Brun-Laguna and Dominique Barthel. Spring 2017.

### *10.2.2. Supervision*

- PhD : Fouzi Boukhalfa, Low and high reliability access in Vehicular Ad-hoc NETworks. Sorbonne University. **Paul Muhlethaler**.
- PhD (ongoing) Mina Rady, Heterogeneous architectures for the IoT, Sorbonne University. **Thomas Watteyne** and **Paul Muhlethaler**, under a CIFRE agreement with Orange Labs, Meylan, France.
- PhD (viva on 18 December 2018) Keoma Brun-Laguna, Deterministic networking for the Industrial Internet of Things, Sorbonne University. **Thomas Watteyne** and **Pascale Minet**.
- PhD (in progress) Jonathan Munoz, Time slotted systems for long range communications, Sorbonne University, **Thomas Watteyne** and **Paul Muhlethaler**.
- PhD (in progress) Amar Abane, Name Data Networks in the Internet of Things, CNAM. Samia Bouzefrane and **Paul Muhlethaler**.
- PhD (in progress) Abdallah Soheby, Etude et evaluation de la dissemination des informations dans la 5G. Eric Renault and **Paul Muhlethaler**.
- PhD (in progress) Mamoudou Sangara, Utilisation de techniques de Machine Leraning dans les reseaux VANETs. Samia Bouzefrane and **Paul Muhlethaler**.
- PhD (in progress) Iman Hmedoush, Connection protocols for the 5G IoT. Cedric Adjih and **Paul Muhlethaler**.

### *10.2.3. Juries*

- HdR:
  - Oyunchimeg Shagdar, "Optimizing wireless communications in dense mobile environments", University of Paris-Saclay, prepared at the University of Versailles-Saint Quentin, September 2018, **Pascale Minet** examiner.
  - Imen Jemili, "Optimisation des ressources dans les réseaux mobiles sans fil", University of Bordeaux, November 2018, **Pascale Minet** examiner.
  - Khaled Boussetta "Dimensionnement, adaptation et placement de fonctionnalités réseaux pour des applications sensibles aux délais", University of Paris 13, December 2018, **Paul Muhlethaler** reviewer.
- PhD:
  - Keoma Brun-Laguna, "Deterministic networking for the Industrial Internet of Things", Sorbonne University, December 2018, **Thomas Watteyne** and **Pascale Minet**, PhD advisers.
  - of Jetmir Haxhibeqiri, "Flexible and Scalable Wireless Communication Solutions for Warehouse Applications", imec – IDLab Ghent – Ghent University, Belgium, December 2018, **Thomas Watteyne** reviewer.
  - Narjes Boulila "Communication vehiculaire garantissant les contraintes temps reel pour les systèmes de prevention de collision", University de la Manouba, December 2018, **Paul Muhlethaler** reviewer.
  - Dinh Van Nguyen "Reseaux de capteurs sans-fil pour la cartographie à l'interieure et la localisation precise servant la navigation à basse vitesse dans les villes intelligentes", Mines ParisTech, December 2018, **Paul Muhlethaler** examiner.
  - Veronica Quintuna Rodriguez "Performance Analysis of Virtualized Network", University of Paris 6, October 2018, **Paul Muhlethaler** examiner.
  - Mohamed Tahar Hammi, "Sécurisation de l'Internet des Objets", Telecom ParisTech, September 2018, **Pascale Minet** examiner.
  - Lilia Lassouaoui, "Ordonnancement et routage pour l'augmentation de la duree de vie dans les reseaux de capteurs sans fil", CNAM, July 2018, **Pascale Minet** examiner.
  - Thiwiza Bellache. "Controle dynamique des communications dans une environnement V2V et V2I", University Versailles Saint Quentin February 2018. **Paul Muhlethaler** examiner.
  - Philippe Erzan "Optimisation de la topologie des reseaux sans fil", University Centrale-Supelec, January 2018, **Paul Muhlethaler** examiner.
  - Moussa Déthié Sarr, "Spécification d'un mécanisme de construction automatique de topologies et d'adressage permettant la gestion dynamique de réseaux de capteurs sans fil lineéaires", University of Clermont, January 2018, **Pascale Minet** reviewer.
  - Pengwenlong Gu, "Détection des comportements malveillants dans les réseaux véhiculaires", Telecom ParisTech, February 2018, **Pascale Minet** examiner.

## 10.3. Popularization

### *10.3.1. Interventions*

- Keoma Brun-Laguna
  - "Solsystem: IoT Industriel et déploiement en milieux réels" démo at RII meeting ("Mobilités et environnements, 20 Novembre 2018, Station F, Paris")
- **Thomas Watteyne**

- The Inria-Paris OpenTestbed (you might already have noticed...). Inria Seminar, Paris, 2 October 2018.
- "Le tour du monde en 1,000 capteurs!". Inria Rocquencourt seminar series. 6 June 2018.
- Construire son Reseau IoT en 10min. Conférence d'ouverture de la semaine des Mathématiques. Inria-Paris, Paris, France, 12 March 2018.

### *10.3.2. Dans les Medias*

- **Thomas Watteyne**
  - Edge computing: on vous dit tout sur les technos qui mettent le cloud dans votre poche, 01net, 10 September 2018.
  - Qu'est-ce que l'IoT? Boston Consultancy Group podcast, 9 September 2018.
  - Retour Semaine des maths 2018 : conference de Thomas Watteyne à l'Inria, Academie de Paris, 15 March 2018.

### *10.3.3. Creation of media or tools for science outreach*

A video concerning the research on protocols for VANETs of Mohamed Hadded has been done by Inria.

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] K. BRUN-LAGUNA.*Deterministic Networking for the Industrial IoT*, Sorbonne Université, December 2018, https://hal.archives-ouvertes.fr/tel-01984357

### Articles in International Peer-Reviewed Journal

[2] I. AMDOUNI, C. ADJIH, N. AITSAADI, P. MUHLETHALER.*Extensive Experimentations on Opportunistic Routing in Wireless Sensor Networks*, in "Sensors", September 2018, vol. 18, n$^{o}$ 9, 3031, https://hal.inria.fr/hal-01961054

[3] C. BORMANN, T. WATTEYNE.*Virtual reassembly buffers in 6LoWPAN - draft-ietf-lwig-6lowpan-virtual-reassembly-00*, in "Internet Engineering Task Force", July 2018, https://hal.inria.fr/hal-01968654

[4] K. BRUN-LAGUNA, A. L. DIEDRICHS, D. DUJOVNE, C. TAFFERNABERRY, R. LEONE, X. VILAJOSANA, T. WATTEYNE.*Using SmartMesh IP in Smart Agriculture and Smart Building applications*, in "Computer Communications", May 2018, vol. 121, p. 83-90 [*DOI :* 10.1016/J.COMCOM.2018.03.010], https://hal.inria.fr/hal-01737060

[5] K. BRUN-LAGUNA, P. H. GOMES, T. WATTEYNE, P. MINET.*Moving Beyond Testbeds? Lessons (We) Learned about Connectivity*, in "IEEE Pervasive Computing", December 2018, https://hal.inria.fr/hal-01968647

[6] T. CHANG, T. WATTEYNE, X. VILAJOSANA, P. H. GOMES.*Constructive Interference in 802.15.4: A Tutorial*, in "Communications Surveys and Tutorials, IEEE Communications Society", December 2018, https://hal.inria.fr/hal-01968646

[7] A. L. DIEDRICHS, F. BROMBERG, D. DUJOVNE, K. BRUN-LAGUNA, T. WATTEYNE.*Prediction of Frost Events using Bayesian networks and Random Forest*, in "IEEE internet of things journal", August 2018 [*DOI :* 10.1109/JIOT.2018.2867333], https://hal.archives-ouvertes.fr/hal-01867780

[8] M. ESTEBAN, D. GLENN, M. VUCINIC, S. LATRÉ, J. FAMAEY, Y. TANAKA, K. BRUN-LAGUNA, X. VILAJOSANA, K. MURAOKA, T. WATTEYNE.*Simulating 6TiSCH Networks*, in "Transactions on emerging telecommunications technologies", August 2018, https://hal.inria.fr/hal-01838566

[9] H. IDOUDI, O. MABROUK, P. MINET, L. A. SAIDANE.*Cluster-based Scheduling for Cognitive Radio Sensor Networks*, in "Journal of Ambient Intelligence and Humanized Computing", January 2018, https://hal.archives-ouvertes.fr/hal-01870909

[10] I. KHOUFI, P. MINET.*An Enhanced Deterministic Beacon Advertising Algorithm for Building TSCH Networks*, in "Annals of Telecommunications - annales des télécommunications", May 2018, https://hal.archives-ouvertes.fr/hal-01870277

[11] P. MINET, I. KHOUFI, B. RMILI.*Beacon Advertising in an IEEE 802.15.4e TSCH Network for Space Launch Vehicles*, in "Acta Astronautica", July 2018, https://hal.archives-ouvertes.fr/hal-01870287

[12] P. MINET, P. MUHLETHALER, I. KHOUFI.*Collision avoidance in shared slots in wireless devices of the Internet of Things: models and simulations*, in "Annals of Telecommunications - annales des télécommunications", December 2018 [*DOI :* 10.1007/s12243-018-0693-9], https://hal.archives-ouvertes.fr/hal-01957245

[13] J. MUNOZ, T. CHANG, X. VILAJOSANA, T. WATTEYNE.*Evaluation of IEEE802.15.4g for Environmental Observations*, in "Sensors", October 2018, https://hal.inria.fr/hal-01968648

[14] A. SOBEHY, E. RENAULT, P. MUHLETHALER.*Position certainty propagation : a localization service for ad-hoc networks*, in "Computers", March 2019, vol. 8, n$^{o}$ 1, p. 6-1 - 6-16 [*DOI :* 10.3390/COMPUTERS8010006], https://hal.archives-ouvertes.fr/hal-01997474

[15] X. VILAJOSANA, B. MARTINEZ, I. VILAJOSANA, T. WATTEYNE.*On the Suitability of 6TiSCH for Wireless Seismic Data Streaming*, in "Internet Technology Letters", January 2018, https://hal.inria.fr/hal-01651949

[16] M. VUCINIC, T. WATTEYNE, X. VILAJOSANA.*Broadcasting Strategies in 6TiSCH Networks*, in "Internet Technology Letters", January 2018, https://hal.inria.fr/hal-01630316

[17] Q. WANG, X. VILAJOSANA, T. WATTEYNE.*6TiSCH Operation Sublayer (6top) Protocol (6P) - RFC8480*, in "Internet Engineering Task Force RFC series", November 2018, https://hal.inria.fr/hal-01968655

[18] T. WATTEYNE, C. BORMANN, P. THUBERT.*LLN Minimal Fragment Forwarding - draft-ietf-6lo-minimal-fragment-00*, in "Internet Engineering Task Force", October 2018, https://hal.inria.fr/hal-01968653

## Invited Conferences

[19] C. A. BOANO, S. DUQUENNOY, A. FÖRSTER, O. GNAWALI, R. JACOB, H.-S. KIM, O. LANDSIEDEL, R. MARFIEVICI, L. MOTTOLA, G. P. PICCO, X. VILAJOSANA, T. WATTEYNE, M. ZIMMERLING.*IoTBench: Towards a Benchmark for Low-power Wireless Networking*, in "CPSBench 2018 - 1st Workshop on Benchmarking Cyber-Physical Networks and Systems", Porto, Portugal, April 2018, https://hal.inria.fr/hal-01968659

[20] M.-R. PALATTELLA, F. SISMONDI, T. CHANG, L. BARON, M. VUCINIC, P. MODERNELL, X. VILAJOSANA, T. WATTEYNE.*F-Interop Platform and Tools: Validating IoT Implementations Faster*, in "AdHoc-Now 2018 - 17th International Conference on Ad Hoc Networks and Wireless", Saint Malo, France, September 2018, p. 1-12, https://hal.inria.fr/hal-01858004

[21] F. RINCON, Y. TANAKA, T. WATTEYNE.*On the Impact of WiFi on 2.4 GHz Industrial IoT Networks*, in "IEEE ICII 2018 - IEEE International Conference on Industrial Internet", Bellevue, WA, United States, October 2018, https://hal.inria.fr/hal-01968656

### International Conferences with Proceedings

[22] T. CHANG, T. WATTEYNE, X. VILAJOSANA.*Competition: OpenWSN, a Development Environment for 6TiSCH*, in "International Conference on Embedded Wireless Systems and Networks (EWSN)", Beijing, China, February 2019, https://hal.inria.fr/hal-01968644

[23] P. JACQUET, M. L. LAMALI, F. MATHIEU.*Collecter un nombre inconnu de coupons*, 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.inria.fr/hal-01787252

[24] P. MINET, E. RENAULT, I. KHOUFI, S. BOUMERDASSI.*Analyzing Traces from a Google Data Center*, in "IWCMC 2018 - 14th International Wireless Communications and Mobile Computing Conference", Limassol, Cyprus, Proceedings of the IWCMC 2018 - 14th International Wireless Communications and Mobile Computing Conference, IEEE Computer Society, June 2018, p. 1167 - 1172 [*DOI :* 10.1109/IWCMC.2018.8450304], https://hal.archives-ouvertes.fr/hal-01870216

[25] P. MINET, E. RENAULT, I. KHOUFI, S. BOUMERDASSI.*Data analysis of a Google data center*, in "CCGRID 2018 : 18th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing", Washington Dc, United States, IEEE Computer Society, May 2018, p. 342 - 343 [*DOI :* 10.1109/CCGRID.2018.00049], https://hal.archives-ouvertes.fr/hal-01867409

[26] P. MINET, Z. SOUA, I. KHOUFI.*An Adaptive Schedule for TSCH Networks in the Industry 4.0*, in "PEMWN 2018 - 7th IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks", Toulouse, France, Proceedings of the PEMWN 2018 - 7th IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks, September 2018, p. 1-6, https://hal.archives-ouvertes.fr/hal-01870252

[27] J. MUNOZ, E. RIOU, X. VILAJOSANA, P. MUHLETHALER, T. WATTEYNE.*Overview of IEEE802.15.4g OFDM and its Applicability to Smart Building Applications*, in "Wireless Days Conference (WD)", Dubai, United Arab Emirates, IEEE, April 2018 [*DOI :* 10.1109/WD.2018.8361707], https://hal.inria.fr/hal-01718175

[28] A. SOBEHY, E. RENAULT, P. MUHLETHALER.*Position Certainty Propagation: A location service for MANETs*, in "MSPN 2018 - 4th International Conference on Mobile, Secure and Programmable Networking", Paris, France, Springer, June 2018, p. 131-142 [*DOI :* 10.1007/978-3-030-03101-5_13], https://hal.archives-ouvertes.fr/hal-01864565

### Conferences without Proceedings

[29] A. ABANE, P. MUHLETHALER, S. BOUZEFRANE, M. DAOUI, A. BATTOU.*Towards evaluating Named Data Networking for the IoT: A framework for OMNeT++*, in "OMNeT Community Summit 2018", Pise, Italy, September 2018, https://hal.archives-ouvertes.fr/hal-01864541

[30] M. HADDED, P. MUHLETHALER, A. LAOUITI.*TDMA scheduling strategies for vehicular ad hoc networks: from a distributed to a centralized approach*, in "SoftCOM 2018 - 26th International Con-

ference on Software, Telecommunications and Computer Networks", Split, Croatia, September 2018 [*DOI :* 10.23919/SOFTCOM.2018.8555781], https://hal.archives-ouvertes.fr/hal-01864550

[31] S. INGRACHEN, N. ACHIR, P. MUHLETHALER, T. DJAMAH, A. BERQIA.*A Collaborative Environment Perception Approach for Vehicular Ad hoc Networks*, in "VTC 2018 - Fall - 2018 IEEE 88th Vehicular Technology Conference", Chicago, United States, August 2018, https://hal.archives-ouvertes.fr/hal-01864608

[32] J. MUNOZ, P. MUHLETHALER, X. VILAJOSANA, T. WATTEYNE.*Why Channel Hopping Makes Sense, even with IEEE802.15.4 OFDM at 2.4 GHz*, in "Global IoT Summit (GIoTS)", Bilbao, Spain, June 2018, https://hal.inria.fr/hal-01756523

[33] L. SALAUN, A. ALLOUM, P. JACQUET.*Adaptive Multiplicity Codes based PIR Protocol for Multi-Cloud Plateform Services*, in "IEEE 5G 2018 - World Forum : Workshop on 5G Cloud Native Design", Santa Clara, California, United States, July 2018, p. 1-8, https://hal.archives-ouvertes.fr/hal-01831322

[34] M. SANGARE, S. BANERJEE, P. MUHLETHALER, S. BOUZEFRANE.*Predicting transmission success with Machine-Learning and Support Vector Machine in VANETs*, in "PEMWN 2018 - 7th IFIP/IEEE International Conference on Performance Evaluation and Modeling in Wired and Wireless Networks", Toulouse, France, September 2018, https://hal.archives-ouvertes.fr/hal-01864671

[35] M. SANGARE, S. BANERJEE, P. MUHLETHALER, S. BOUZEFRANE.*Predicting Vehicles' Positions using Roadside Units: a Machine-Learning Approach*, in "IEEE CSCN 2018- IEEE Conference on Standards for Communications and Networking", Paris, France, October 2018, https://hal.archives-ouvertes.fr/hal-01876180

[36] C. Y. TAZIBT, N. ACHIR, P. MUHLETHALER, T. DJAMAH.*UAV-based Data Gathering using An Artificial Potential Fields Approach*, in "VTC 2018-Fall - IEEE 88th Vehicular Technology Conference", Chicago, United States, August 2018, https://hal.archives-ouvertes.fr/hal-01864590

[37] M. VUCINIC, M. PEJANOVIC-DJURISIC, T. WATTEYNE.*SODA: 6TiSCH Open Data Action*, in "CPSBench 2018 - International Workshop on Benchmarking Cyber-Physical Networks and Systems", Porto, Portugal, April 2018, https://hal.inria.fr/hal-01730483

## Research Reports

[38] M. FERREIRA, J. MUNOZ, T. WATTEYNE.*SmartMesh Range Measurements*, Inria, September 2018, n$^{o}$ RR-9205, 18, https://hal.inria.fr/hal-01874919

## Other Publications

[39] B. BŁASZCZYSZYN, P. MUHLETHALER.*Analyzing LoRa long-range, low-power, wide-area networks using stochastic geometry*, December 2018, https://arxiv.org/abs/1811.01886 - 8 pages, https://hal.inria.fr/hal-01958939

[40] P. JACQUET, W. SZPANKOWSKI.*Distribution of Tail Symbols in DST for Markov Sources*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01966158

# Project-Team GALLIUM

## Programming languages, types, compilation and proofs

RESEARCH CENTER
**Paris**

THEME
**Proofs and Verification**

# Table of contents

# Project-Team GALLIUM

*Creation of the Project-Team: 2006 May 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.1.1. - Multicore, Manycore
A1.1.3. - Memory models
A2.1. - Programming Languages
A2.1.1. - Semantics of programming languages
A2.1.3. - Object-oriented programming
A2.1.4. - Functional programming
A2.1.6. - Concurrent programming
A2.1.11. - Proof languages
A2.2. - Compilation
A2.2.1. - Static analysis
A2.2.2. - Memory models
A2.2.4. - Parallel architectures
A2.2.5. - Run-time systems
A2.4. - Formal method for verification, reliability, certification
A2.4.1. - Analysis
A2.4.3. - Proofs
A2.5.4. - Software Maintenance & Evolution
A7.1.2. - Parallel algorithms
A7.2. - Logic in Computer Science
A7.2.2. - Automated Theorem Proving
A7.2.3. - Interactive Theorem Proving

**Other Research Topics and Application Domains:**

B5.2.3. - Aviation
B6.1. - Software industry
B6.6. - Embedded systems
B9.5.1. - Computer science

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Xavier Leroy [Team leader, Inria, Senior Researcher, until Oct 2018; Collège de France, Prof, since Nov 2018]
Umut Acar [Inria, Advanced Research Position, until Apr 2018]
Damien Doligez [Inria, Researcher]
Ioannis Filippidis [Inria, Starting Research Position, since Oct 2018]
Fabrice Le Fessant [Inria, Researcher, until Jan 2018]
Jean-Marie Madiot [Inria, Researcher]
Luc Maranget [Inria, Researcher]
Michel Mauny [Inria, Senior Researcher]

François Pottier [Inria, Senior Researcher, HDR]
Mike Rainey [Inria, Starting Research Position, until Feb 2018]
Didier Rémy [Inria, Senior Researcher, HDR]

**Technical Staff**
Sébastien Hinderer [Inria]

**PhD Students**
Vitaly Aksenov [Inria, until Aug 2018]
Armaël Guéneau [Université Paris Diderot]
Glen Mével [Inria, since Nov 2018]
Naomi Testard [Inria]
Thomas Williams [ENS Paris, until Aug 2018]

**Post-Doctoral Fellow**
Gergö Barany [Inria, until Aug 2018]

**Administrative Assistant**
Laurence Bourcier [Inria]

# 2. Overall Objectives

## 2.1. Research at Gallium

The research conducted in the Gallium group aims at improving the safety, reliability and security of software through advances in programming languages and formal verification of programs. Our work is centered on the design, formalization and implementation of functional programming languages, with particular emphasis on type systems and type inference, formal verification of compilers, and interactions between programming and program proof. The OCaml language and the CompCert verified C compiler embody many of our research results. Our work spans the whole spectrum from theoretical foundations and formal semantics to applications to real-world problems.

# 3. Research Program

## 3.1. Programming languages: design, formalization, implementation

Like all languages, programming languages are the media by which thoughts (software designs) are communicated (development), acted upon (program execution), and reasoned upon (validation). The choice of adequate programming languages has a tremendous impact on software quality. By "adequate", we mean in particular the following four aspects of programming languages:

- **Safety.** The programming language must not expose error-prone low-level operations (explicit memory deallocation, unchecked array access, etc) to programmers. Further, it should provide constructs for describing data structures, inserting assertions, and expressing invariants within programs. The consistency of these declarations and assertions should be verified through compile-time verification (e.g. static type-checking) and run-time checks.
- **Expressiveness.** A programming language should manipulate as directly as possible the concepts and entities of the application domain. In particular, complex, manual encodings of domain notions into programmatic notations should be avoided as much as possible. A typical example of a language feature that increases expressiveness is pattern matching for examination of structured data (as in symbolic programming) and of semi-structured data (as in XML processing). Carried to the extreme, the search for expressiveness leads to domain-specific languages, customized for a specific application area.

- **Modularity and compositionality.** The complexity of large software systems makes it impossible to design and develop them as one, monolithic program. Software decomposition (into semi-independent components) and software composition (of existing or independently-developed components) are therefore crucial. Again, this modular approach can be applied to any programming language, given sufficient fortitude by the programmers, but is much facilitated by adequate linguistic support. In particular, reflecting notions of modularity and software components in the programming language enables compile-time checking of correctness conditions such as type correctness at component boundaries.
- **Formal semantics.** A programming language should fully and formally specify the behaviours of programs using mathematical semantics, as opposed to informal, natural-language specifications. Such a formal semantics is required in order to apply formal methods (program proof, model checking) to programs.

Our research work in language design and implementation centers on the statically-typed functional programming paradigm, which scores high on safety, expressiveness and formal semantics, complemented with full imperative features and objects for additional expressiveness, and modules and classes for compositionality. The OCaml language and system embodies many of our earlier results in this area [37]. Through collaborations, we also gained experience with several domain-specific languages based on a functional core, including distributed programming (JoCaml), XML processing (XDuce, CDuce), reactive functional programming, and hardware modeling.

## 3.2. Type systems

Type systems [39] are a very effective way to improve programming language reliability. By grouping the data manipulated by the program into classes called types, and ensuring that operations are never applied to types over which they are not defined (e.g. accessing an integer as if it were an array, or calling a string as if it were a function), a tremendous number of programming errors can be detected and avoided, ranging from the trivial (misspelled identifier) to the fairly subtle (violation of data structure invariants). These restrictions are also very effective at thwarting basic attacks on security vulnerabilities such as buffer overflows.

The enforcement of such typing restrictions is called type-checking, and can be performed either dynamically (through run-time type tests) or statically (at compile-time, through static program analysis). We favor static type-checking, as it catches bugs earlier and even in rarely-executed parts of the program, but note that not all type constraints can be checked statically if static type-checking is to remain decidable (i.e. not degenerate into full program proof). Therefore, all typed languages combine static and dynamic type-checking in various proportions.

Static type-checking amounts to an automatic proof of partial correctness of the programs that pass the compiler. The two key words here are *partial*, since only type safety guarantees are established, not full correctness; and *automatic*, since the proof is performed entirely by machine, without manual assistance from the programmer (beyond a few, easy type declarations in the source). Static type-checking can therefore be viewed as the poor man's formal methods: the guarantees it gives are much weaker than full formal verification, but it is much more acceptable to the general population of programmers.

### *3.2.1. Type systems and language design.*

Unlike most other uses of static program analysis, static type-checking rejects programs that it cannot prove safe. Consequently, the type system is an integral part of the language design, as it determines which programs are acceptable and which are not. Modern typed languages go one step further: most of the language design is determined by the *type structure* (type algebra and typing rules) of the language and intended application area. This is apparent, for instance, in the XDuce and CDuce domain-specific languages for XML transformations [35], [32], whose design is driven by the idea of regular expression types that enforce DTDs at compile-time. For this reason, research on type systems – their design, their proof of semantic correctness (type safety), the development and proof of associated type-checking and inference algorithms – plays a large and central role in the field of programming language research, as evidenced by the huge number of type systems papers in conferences such as Principles of Programming Languages.

### *3.2.2. Polymorphism in type systems.*

There exists a fundamental tension in the field of type systems that drives much of the research in this area. On the one hand, the desire to catch as many programming errors as possible leads to type systems that reject more programs, by enforcing fine distinctions between related data structures (say, sorted arrays and general arrays). The downside is that code reuse becomes harder: conceptually identical operations must be implemented several times (say, copying a general array and a sorted array). On the other hand, the desire to support code reuse and to increase expressiveness leads to type systems that accept more programs, by assigning a common type to broadly similar objects (for instance, the `Object` type of all class instances in Java). The downside is a loss of precision in static typing, requiring more dynamic type checks (downcasts in Java) and catching fewer bugs at compile-time.

*Polymorphic* type systems offer a way out of this dilemma by combining precise, descriptive types (to catch more errors statically) with the ability to abstract over their differences in pieces of reusable, generic code that is concerned only with their commonalities. The paradigmatic example is parametric polymorphism, which is at the heart of all typed functional programming languages. Many forms of polymorphic typing have been studied since then. Taking examples from our group, the work of Rémy, Vouillon and Garrigue on row polymorphism [42], integrated in OCaml, extended the benefits of this approach (reusable code with no loss of typing precision) to object-oriented programming, extensible records and extensible variants. Another example is the work by Pottier on subtype polymorphism, using a constraint-based formulation of the type system [40]. Finally, the notion of "coercion polymorphism" proposed by Cretin and Rémy[5] combines and generalizes both parametric and subtyping polymorphism.

### *3.2.3. Type inference.*

Another crucial issue in type systems research is the issue of type inference: how many type annotations must be provided by the programmer, and how many can be inferred (reconstructed) automatically by the type-checker? Too many annotations make the language more verbose and bother the programmer with unnecessary details. Too few annotations make type-checking undecidable, possibly requiring heuristics, which is unsatisfactory. OCaml requires explicit type information at data type declarations and at component interfaces, but infers all other types.

In order to be predictable, a type inference algorithm must be complete. That is, it must not find *one*, but *all* ways of filling in the missing type annotations to form an explicitly typed program. This task is made easier when all possible solutions to a type inference problem are *instances* of a single, *principal* solution.

Maybe surprisingly, the strong requirements – such as the existence of principal types – that are imposed on type systems by the desire to perform type inference sometimes lead to better designs. An illustration of this is row variables. The development of row variables was prompted by type inference for operations on records. Indeed, previous approaches were based on subtyping and did not easily support type inference. Row variables have proved simpler than structural subtyping and more adequate for type-checking record update, record extension, and objects.

Type inference encourages abstraction and code reuse. A programmer's understanding of his own program is often initially limited to a particular context, where types are more specific than strictly required. Type inference can reveal the additional generality, which allows making the code more abstract and thus more reuseable.

## 3.3. Compilation

Compilation is the automatic translation of high-level programming languages, understandable by humans, to lower-level languages, often executable directly by hardware. It is an essential step in the efficient execution, and therefore in the adoption, of high-level languages. Compilation is at the interface between programming languages and computer architecture, and because of this position has had considerable influence on the design of both. Compilers have also attracted considerable research interest as the oldest instance of symbolic processing on computers.

Compilation has been the topic of much research work in the last 40 years, focusing mostly on high-performance execution ("optimization") of low-level languages such as Fortran and C. Two major results came out of these efforts: one is a superb body of performance optimization algorithms, techniques and methodologies; the other is the whole field of static program analysis, which now serves not only to increase performance but also to increase reliability, through automatic detection of bugs and establishment of safety properties. The work on compilation carried out in the Gallium group focuses on a less investigated topic: compiler certification.

### *3.3.1. Formal verification of compiler correctness.*

While the algorithmic aspects of compilation (termination and complexity) have been well studied, its semantic correctness – the fact that the compiler preserves the meaning of programs – is generally taken for granted. In other terms, the correctness of compilers is generally established only through testing. This is adequate for compiling low-assurance software, themselves validated only by testing: what is tested is the executable code produced by the compiler, therefore compiler bugs are detected along with application bugs. This is not adequate for high-assurance, critical software which must be validated using formal methods: what is formally verified is the source code of the application; bugs in the compiler used to turn the source into the final executable can invalidate the guarantees so painfully obtained by formal verification of the source.

To establish strong guarantees that the compiler can be trusted not to change the behavior of the program, it is necessary to apply formal methods to the compiler itself. Several approaches in this direction have been investigated, including translation validation, proof-carrying code, and type-preserving compilation. The approach that we currently investigate, called *compiler verification*, applies program proof techniques to the compiler itself, seen as a program in particular, and use a theorem prover (the Coq system) to prove that the generated code is observationally equivalent to the source code. Besides its potential impact on the critical software industry, this line of work is also scientifically fertile: it improves our semantic understanding of compiler intermediate languages, static analyses and code transformations.

## 3.4. Interface with formal methods

Formal methods collectively refer to the mathematical specification of software or hardware systems and to the verification of these systems against these specifications using computer assistance: model checkers, theorem provers, program analyzers, etc. Despite their costs, formal methods are gaining acceptance in the critical software industry, as they are the only way to reach the required levels of software assurance.

In contrast with several other Inria projects, our research objectives are not fully centered around formal methods. However, our research intersects formal methods in the following two areas, mostly related to program proofs using proof assistants and theorem provers.

### *3.4.1. Software-proof codesign*

The current industrial practice is to write programs first, then formally verify them later, often at huge costs. In contrast, we advocate a codesign approach where the program and its proof of correctness are developed in interaction, and we are interested in developing ways and means to facilitate this approach. One possibility that we currently investigate is to extend functional programming languages such as OCaml with the ability to state logical invariants over data structures and pre- and post-conditions over functions, and interface with automatic or interactive provers to verify that these specifications are satisfied. Another approach that we practice is to start with a proof assistant such as Coq and improve its capabilities for programming directly within Coq.

### *3.4.2. Mechanized specifications and proofs for programming language components*

We emphasize mathematical specifications and proofs of correctness for key language components such as semantics, type systems, type inference algorithms, compilers and static analyzers. These components are getting so large that machine assistance becomes necessary to conduct these mathematical investigations. We have already mentioned using proof assistants to verify compiler correctness. We are also interested in

using them to specify and reason about semantics and type systems. These efforts are part of a more general research topic that is gaining importance: the formal verification of the tools that participate in the construction and certification of high-assurance software.

# 4. Application Domains

## 4.1. High-assurance software

A large part of our work on programming languages and tools focuses on improving the reliability of software. Functional programming, program proof, and static type-checking contribute significantly to this goal.

Because of its proximity with mathematical specifications, pure functional programming is well suited to program proof. Moreover, functional programming languages such as OCaml are eminently suitable to develop the code generators and verification tools that participate in the construction and qualification of high-assurance software. Examples include Esterel Technologies's KCG 6 code generator, the Astrée static analyzer, the Caduceus/Jessie program prover, and the Frama-C platform. Our own work on compiler verification combines these two aspects of functional programming: writing a compiler in a pure functional language and mechanically proving its correctness.

Static typing detects programming errors early, prevents a number of common sources of program crashes (null dereferences, out-of bound array accesses, etc), and helps tremendously to enforce the integrity of data structures. Judicious uses of generalized abstract data types (GADTs), phantom types, type abstraction and other encapsulation mechanisms also allow static type checking to enforce program invariants.

## 4.2. Software security

Static typing is also highly effective at preventing a number of common security attacks, such as buffer overflows, stack smashing, and executing network data as if it were code. Applications developed in a language such as OCaml are therefore inherently more secure than those developed in unsafe languages such as C.

The methods used in designing type systems and establishing their soundness can also deliver static analyses that automatically verify some security policies. Two examples from our past work include Java bytecode verification [38] and enforcement of data confidentiality through type-based inference of information flow and noninterference properties [41].

## 4.3. Processing of complex structured data

Like most functional languages, OCaml is very well suited to expressing processing and transformations of complex, structured data. It provides concise, high-level declarations for data structures; a very expressive pattern-matching mechanism to destructure data; and compile-time exhaustiveness tests. Therefore, OCaml is an excellent match for applications involving significant amounts of symbolic processing: compilers, program analyzers and theorem provers, but also (and less obviously) distributed collaborative applications, advanced Web applications, financial modeling tools, etc.

## 4.4. Rapid development

Static typing is often criticized as being verbose (due to the additional type declarations required) and inflexible (due to, for instance, class hierarchies that must be fixed in advance). Its combination with type inference, as in the OCaml language, substantially diminishes the importance of these problems: type inference allows programs to be initially written with few or no type declarations; moreover, the OCaml approach to object-oriented programming completely separates the class inheritance hierarchy from the type compatibility relation. Therefore, the OCaml language is highly suitable for fast prototyping and the gradual evolution of software prototypes into final applications, as advocated by the popular "extreme programming" methodology.

## 4.5. Teaching programming

Our work on the OCaml language family has an impact on the teaching of programming. OCaml is one of the programming languages selected by the French Ministry of Education for teaching Computer Science in *classes préparatoires scientifiques*. OCaml is also widely used for teaching advanced programming in engineering schools, colleges and universities in France, the USA, and Japan.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

In 2018, Xavier Leroy received the "Grand prix" jointly awarded by Inria and Académie des sciences.

Gergö Barany received the Best Paper Award for the paper "Finding Missed Compiler Optimizations by Differential Testing" [19] at the 27th International Conference on Compiler Construction (CC 2018).

# 6. New Software and Platforms

## 6.1. Compcert

*The CompCert formally-verified C compiler*
KEYWORDS: Compilers - Formal methods - Deductive program verification - C - Coq
FUNCTIONAL DESCRIPTION: CompCert is a compiler for the C programming language. Its intended use is the compilation of life-critical and mission-critical software written in C and meeting high levels of assurance. It accepts most of the ISO C 99 language, with some exceptions and a few extensions. It produces machine code for the ARM, PowerPC, RISC-V, and x86 architectures. What sets CompCert C apart from any other production compiler, is that it is formally verified to be exempt from miscompilation issues, using machine-assisted mathematical proofs (the Coq proof assistant). In other words, the executable code it produces is proved to behave exactly as specified by the semantics of the source C program. This level of confidence in the correctness of the compilation process is unprecedented and contributes to meeting the highest levels of software assurance. In particular, using the CompCert C compiler is a natural complement to applying formal verification techniques (static analysis, program proof, model checking) at the source code level: the correctness proof of CompCert C guarantees that all safety properties verified on the source code automatically hold as well for the generated executable.

RELEASE FUNCTIONAL DESCRIPTION: Novelties include a formally-verified type checker for CompCert C, a more careful modeling of pointer comparisons against the null pointer, algorithmic improvements in the handling of deeply nested struct and union types, much better ABI compatibility for passing composite values, support for GCC-style extended inline asm, and more complete generation of DWARF debugging information (contributed by AbsInt).

- Participants: Xavier Leroy, Sandrine Blazy, Jacques-Henri Jourdan, Sylvie Boldo and Guillaume Melquiond
- Partner: AbsInt Angewandte Informatik GmbH
- Contact: Xavier Leroy
- URL: http://compcert.inria.fr/

## 6.2. Diy

*Do It Yourself*
KEYWORD: Parallelism

FUNCTIONAL DESCRIPTION: The diy suite provides a set of tools for testing shared memory models: the litmus tool for running tests on hardware, various generators for producing tests from concise specifications, and herd, a memory model simulator. Tests are small programs written in x86, Power or ARM assembler that can thus be generated from concise specification, run on hardware, or simulated on top of memory models. Test results can be handled and compared using additional tools.

- Participants: Jade Alglave and Luc Maranget
- Partner: University College London UK
- Contact: Luc Maranget
- URL: http://diy.inria.fr/

## 6.3. Menhir

KEYWORDS: Compilation - Context-free grammars - Parsing
FUNCTIONAL DESCRIPTION: Menhir is a LR(1) parser generator for the OCaml programming language. That is, Menhir compiles LR(1) grammar specifications down to OCaml code. Menhir was designed and implemented by François Pottier and Yann Régis-Gianas.

- Contact: François Pottier
- Publications: A Simple, Possibly Correct LR Parser for C11 - Reachability and Error Diagnosis in LR(1) Parsers

## 6.4. OCaml

KEYWORDS: Functional programming - Static typing - Compilation
FUNCTIONAL DESCRIPTION: The OCaml language is a functional programming language that combines safety with expressiveness through the use of a precise and flexible type system with automatic type inference. The OCaml system is a comprehensive implementation of this language, featuring two compilers (a bytecode compiler, for fast prototyping and interactive use, and a native-code compiler producing efficient machine code for x86, ARM, PowerPC and System Z), a debugger, a documentation generator, a compilation manager, a package manager, and many libraries contributed by the user community.

- Participants: Damien Doligez, Xavier Leroy, Fabrice Le Fessant, Luc Maranget, Gabriel Scherer, Alain Frisch, Jacques Garrigue, Marc Shinwell, Jeremy Yallop and Leo White
- Contact: Damien Doligez
- URL: https://ocaml.org/

## 6.5. PASL

KEYWORD: Parallel computing
FUNCTIONAL DESCRIPTION: PASL is a C++ library for writing parallel programs targeting the broadly available multicore computers. The library provides a high level interface and can still guarantee very good efficiency and performance, primarily due to its scheduling and automatic granularity control mechanisms.

- Participants: Arthur Charguéraud, Michael Rainey and Umut Acar
- Contact: Michael Rainey
- URL: http://deepsea.inria.fr/pasl/

## 6.6. ZENON

FUNCTIONAL DESCRIPTION: Zenon is an automatic theorem prover based on the tableaux method. Given a first-order statement as input, it outputs a fully formal proof in the form of a Coq proof script. It has special rules for efficient handling of equality and arbitrary transitive relations. Although still in the prototype stage, it already gives satisfying results on standard automatic-proving benchmarks.

Zenon is designed to be easy to interface with front-end tools (for example integration in an interactive proof assistant), and also to be retargeted to output scripts for different frameworks (for example, Isabelle and Dedukti).

- Author: Damien Doligez
- Contact: Damien Doligez
- URL: http://zenon-prover.org/

## 6.7. OPAM Builder

KEYWORDS: Ocaml - Continuous integration - Opam
FUNCTIONAL DESCRIPTION: OPAM Builder checks in real-time the installability on a computer of all packages after any modification of the repository. To achieve this result, it uses smart mechanisms to compute incremental differencies between package updates, to be able to reuse cached compilations, and switch from a quadratic complexity to a linear complexity.

- Partner: OCamlPro
- Contact: Fabrice Le Fessant
- URL: http://github.com/OCamlPro/opam-builder

## 6.8. TLAPS

*TLA+ proof system*
KEYWORD: Proof assistant
FUNCTIONAL DESCRIPTION: TLAPS is a platform for developing and mechanically verifying proofs about TLA+ specifications. The TLA+ proof language is hierarchical and explicit, allowing a user to decompose the overall proof into proof steps that can be checked independently. TLAPS consists of a proof manager that interprets the proof language and generates a collection of proof obligations that are sent to backend verifiers. The current backends include the tableau-based prover Zenon for first-order logic, Isabelle/TLA+, an encoding of TLA+ set theory as an object logic in the logical framework Isabelle, an SMT backend designed for use with any SMT-lib compatible solver, and an interface to a decision procedure for propositional temporal logic.
NEWS OF THE YEAR: Ioannis Filippidis joined the development team in November 2018 and started designing and implementing support for reasoning about TLA+'s ENABLED construct.

- Participants: Damien Doligez, Stephan Merz and IOANNIS FILIPPIDIS
- Contact: Stephan Merz
- URL: https://tla.msr-inria.inria.fr/tlaps/content/Home.html

## 6.9. CFML

*Interactive program verification using characteristic formulae*
KEYWORDS: Coq - Software Verification - Deductive program verification - Separation Logic
FUNCTIONAL DESCRIPTION: The CFML tool supports the verification of OCaml programs through interactive Coq proofs. CFML proofs establish the full functional correctness of the code with respect to a specification. They may also be used to formally establish bounds on the asymptotic complexity of the code. The tool is made of two parts: on the one hand, a characteristic formula generator implemented as an OCaml program that parses OCaml code and produces Coq formulae, and, on the other hand, a Coq library that provides notations and tactics for manipulating characteristic formulae interactively in Coq.

- Participants: Arthur Charguéraud, Armaël Guéneau and François Pottier
- Contact: Arthur Charguéraud
- URL: http://www.chargueraud.org/softs/cfml/

## 6.10. ldrgen

*Liveness-driven random C code generator*
KEYWORDS: Code generation - Randomized algorithms - Static program analysis
FUNCTIONAL DESCRIPTION: The ldrgen program is a generator of C code: On every call it generates a new random C function and prints it to the standard output. The generator is "liveness-driven", which means that it tries to avoid generating dead code: All the computations it generates are (in a certain, limited sense) actually used to compute the function's return value. This is achieved by generating the program backwards, in combination with a simultaneous liveness analysis that guides the random generator's choices.

- Participant: Gergö Barany
- Contact: Gergö Barany
- Publication: Liveness-Driven Random Program Generation
- URL: https://github.com/gergo-/ldrgen

# 7. New Results

## 7.1. Formal verification of compilers and static analyzers

### *7.1.1. The CompCert formally-verified compiler*

**Participants:** Xavier Leroy, Daniel Kästner [AbsInt GmbH], Michael Schmidt [AbsInt GmbH], Bernhard Schommer [AbsInt GmbH].

In the context of our work on compiler verification (see section 3.3.1), since 2005, we have been developing and formally verifying a moderately-optimizing compiler for a large subset of the C programming language, generating assembly code for the ARM, PowerPC, RISC-V and x86 architectures [9]. This compiler comprises a back-end part, translating the Cminor intermediate language to PowerPC assembly and reusable for source languages other than C [8], and a front-end translating the CompCert C subset of C to Cminor. The compiler is mostly written within the specification language of the Coq proof assistant, from which Coq's extraction facility generates executable OCaml code. The compiler comes with a 100000-line machine-checked Coq proof of semantic preservation establishing that the generated assembly code executes exactly as prescribed by the semantics of the source C program.

This year, we improved the CompCert C compiler in several directions:

- A new built-in function, `__builtin_ais_annot` makes it easy to transfer annotations (also known as flow facts) written at the source code level in AbsInt's aiS annotation language all the way down to the level of the generated machine code. The aiT static analyzer for Worst-Case Execution Times, which operates at the machine code level, can then take advantage of these annotations to produce better WCET estimates.
- In preparation for a qualification with respect to industry standards for avionics software, conformance with the ISO C 1999 and ISO C 2011 standards was improved, with the addition of many diagnostics required by the standards.
- Performance of the generated code was slightly improved via changes to the heuristics for function inlining and for instruction selection.
- The semantic modeling of external function calls was made more precise, reflecting the fact that these functions can destroy some registers and some stack locations.

We released three versions of CompCert incorporating these improvements: version 3.2 in January 2018, version 3.3 in May 2018, and version 3.4 in September 2018.

Two papers on CompCert were presented at conferences. The first paper, with Daniel Kästner as lead author, was presented at the 2018 ERTS congress [22]. It describes the use of CompCert to compile software for nuclear power plant equipment developed by MTU Friedrichshafen, and the required certification of CompCert according to the IEC 60880 regulations for the nuclear industry. The second paper, with Bernhard Schommer as lead author, was presented at the 2018 WCET workshop [23]. It describes the `__builtin_ais_annot` source-level annotation mechanism mentioned above and its uses to help WCET analysis.

### *7.1.2. Verified code generation in the polyhedral model*

**Participants:** Nathanaël Courant, Xavier Leroy.

The polyhedral model is a high-level intermediate representation for loop nests iterating over arrays and matrices, as found in numerical code. It supports a great many loop optimizations (fusion, splitting, interchange, blocking, etc) in a uniform, mathematically-elegant manner.

Nathanaël Courant, as part of his MPRI Master's internship and under Xavier Leroy's supervision, developed a Coq formalization of the polyhedral model. He then implemented and proved correct in Coq a code generator that produces efficient sequential code from an optimized polyhedral representation. Code generation is a delicate part of polyhedral compilation, involving complex, error-prone algorithms. Nathanaël Courant's verified code generator includes the major algorithms from Cédric Bastoul's reference paper [31]. The Coq specifications and proofs are available at https://github.com/Ekdohibs/PolyGen.

### *7.1.3. Testing compiler optimizations*

**Participant:** Gergö Barany.

Compilers should be correct, but they should ideally also generate machine code that is as efficient as possible. Gergö Barany continued work on testing the quality of the generated code.

In a differential testing approach, one generates random C programs, compiles them with different compilers, then compares the generated code using a custom binary analysis tool. This tool finds missed optimizations by comparing criteria such as the number of instructions, the number of reads from the stack (for comparing the quality of register spilling), or the numbers of various other classes of instructions affected by optimizations of interest.

The system has found previously unreported missing optimizations in the GCC, Clang, and CompCert compilers. An article [19] was presented at the 27th International Conference on Compiler Construction (CC 2018), where it was honored with the Best Paper Award.

### *7.1.4. A verified model of register aliasing in CompCert*

**Participants:** Gergö Barany, Xavier Leroy.

Some CPU architectures such as ARM feature register aliasing: Each of its 64-bit floating-point registers can also be accessed as two separate 32-bit halves. Modifying a superregister changes (invalidates) the data stored in subregisters and vice versa, but this behavior was not yet modeled in CompCert's semantics.

We continued work on re-engineering much of CompCert's semantic model of the register file and of the call stack. Rather than simple mappings of locations to values, the register file and the stack are now modeled more realistically as blocks of memory containing bytes that represent fragments of values. In this way, we can verify a semantic model in which a 64-bit register or stack slot may contain either a single 64-bit value or a pair of two unrelated 32-bit values. This ongoing work was presented at the workshop on Syntax and Semantics of Low-Level Languages (LOLA 2018) [25].

## 7.2. Language design and type systems

### *7.2.1. Refactoring with ornaments in ML*

**Participants:** Thomas Williams, Lucas Baudin, Didier Rémy.

Thomas Williams, Lucas Baudin, and Didier Rémy have been working on refactoring and other transformations of ML programs based on mixed ornamentation and disornamentation. Ornaments have been introduced as a way of describing changes in data type definitions that can reorganize or add pieces of data. After a new data structure has been described as an ornament of an older one, the functions that operate on the bare structure can be partially or sometimes totally lifted into functions that operate on the ornamented structure.

Williams and Rémy improved the formalisation of the lifting framework: using ornament inference, an ML program is first elaborated into a generic program, which can be seen as a template for all possible liftings of the original program. The generic program is defined in a superset of ML. It can then be instantiated with specific ornaments, and simplified back to an ML program. Williams and Rémy studied the semantics of this intermediate language and used it to prove the correctness of the lifting, using logical relations techniques. This work has been presented at POPL 2018 [12]. More technical details appear in a research report [43].

Lucas Baudin and Dider Rémy also studied the inverse transformation, disornamentation, which allows removing pieces of information from a data structure and adjusting the code accordingly. They showed that the framework of ornamentation can also be used to allow mixed ornamentation and disornamentation transformations. They also designed a new patch language to describe in a more robust manner how the code must be modified during such transformations. This enables a new class of applications, such as maintaining two views of a data structure in sync. For example, the location information in an abstract syntax tree, which is used to report error messages but obfuscates the code, can be projected away, leading to a simpler version of the code, which can then be modified and often automatically reornamented into the richer version of the code with locations. Disonamentation has been presented by Lucas Baudin at the ML 2018 workshop. Ornamentation, including mixed disornamentation, has also been presented at the MSFP 2018 workshop in Oxford.

A small prototype with ornamentation has been written by Thomas Williams and extended with disornamentation by Lucas Baudin. Thomas Williams has also started developing a new version of the prototype that will handle most of the OCaml language.

## 7.3. Shared-memory concurrency

### *7.3.1. The Linux Kernel Memory Model*

**Participants:** Luc Maranget, Jade Alglave [University College London & ARM Ltd], Paul Mckenney [IBM Corporation], Andrea Parri [Sant'Anna School of Advanced Studies, Pisa, Italy], Alan Stern [Harvard University].

Modern multi-core and multi-processor computers do not follow the intuitive "sequential consistency" model that would define a concurrent execution as the interleaving of the executions of its constituent threads and that would command instantaneous writes to the shared memory. This situation is due both to in-core optimisations such as speculative and out-of-order execution of instructions, and to the presence of sophisticated (and cooperating) caching devices between processors and memory. Luc Maranget is taking part in an international research effort to define the semantics of the computers of the multi-core era, and more generally of shared-memory parallel devices or languages, with a clear initial focus on devices.

This year saw a publication on languages in an international conference. A multi-year effort to define a weak memory model for the Linux Kernel has yielded a scholarly paper [18] presented at the *Architectural Support for Programming Languages and Operating Systems* (ASPLOS) conference in March 2018. The article describes a formal model, the *Linux Kernel Memory Model* (LKMM), which defines how Linux kernel programs are supposed to behave. The model, a CAT model, can be simulated using the **herd** simulator, allowing programmers to experiment and develop intuitions. The model was tested against hardware and refined in consultation with Linux maintainers. Finally, the ASPLOS paper formalizes the *fundamental law of the Read-Copy-Update synchronization mechanism* and proves that one of its implementations satisfies this law. It is worth noting that the LKMM is now part of the Linux kernel source (in the `tools/`) section). Luc Maranget and his co-authors are the official maintainers of this document.

### *7.3.2. The ARMv8 and RISC-V memory model*

**Participants:** Will Deacon [ARM Ltd], Luc Maranget, Jade Alglave [University College London & ARM Ltd].

Jade Alglave and Luc Maranget are working on a mixed-size version of the ARMv8 memory model. This model builds on the `aarch64.cat` model authored last year by Will Deacon (ARM Ltd). This ongoing work is subject to IP restrictions which we hope to lift next year.

Luc Maranget is an individual member of the memory model group of the RISC-V consortium (https://riscv.org/). Version V2.3 of the User-Level ISA Specification is now complete and should be released soon. This version features the first occurrence of a detailed memory model expressed in English, as well as its transliteration in CAT authored by Luc Maranget.

### *7.3.3. Work on diy*

**Participant:** Luc Maranget.

This year, new synchronisation primitives were added to the Linux kernel memory model; ARMv8 atomic instructions were added; and more.

A more significant improvement is the introduction of *mixed-size* accesses. The tools can now handle a new view of memory, where memory is made up of elementary cells (typically *bytes*) that can be read or written as groups of contiguous cells (typically up to *quadwords* of 8 bytes). This preliminary work paves the way to the simulation of more elaborate memory models.

### *7.3.4. Unifying axiomatic and operational weak memory models*

**Participants:** Jean-Marie Madiot, Jade Alglave [University College London & ARM Ltd], Simon Castellan [Imperial College London].

Modern multi-processors optimize the running speed of programs using a variety of techniques, including caching, instruction reordering, and branch speculation. While those techniques are perfectly invisible to sequential programs, such is not the case for concurrent programs that execute several threads and share memory: threads do not share at every point in time a single consistent view of memory. A *weak memory model* offers only weak consistency guarantees when reasoning about the permitted behaviors of a program. Until now, there have been two kinds of such models, based on different mathematical foundations: axiomatic models and operational models.

Axiomatic models explicitly represent the dependencies between the program and memory actions. These models are convenient for causal reasoning about programs. They are also well-suited to the simulation and testing of *hardware* microprocessors.

Operational models represent program states directly, thus can be used to reason on programs: program logics become applicable, and the reasoning behind nondeterministic behavior is much clearer. This makes them preferable for reasoning about *software*.

Jean-Marie Madiot has been collaborating with weak memory model expert Jade Alglave and concurrent game semantics researcher Simon Castellan in order to unify these styles, in a way that attempts to combine the best of both approaches. The first results are a formalisation of TSO-style architectures using partial-order techniques similar to the ones used in game semantics, and a proof of a stronger-than-state-of-art "data-race freedom" theorem: well-synchronised programs can assume a strong memory model. These results have been submitted for publication.

This is a first step towards tractable verification of concurrent programs, combining software verification using concurrent program logics, in the top layer, and hardware testing using weak memory models, in the bottom layer. Our hope is to leave no unverified gap between software and hardware, even (and especially) in the presence of concurrency.

### 7.3.5. Granularity control for parallel programs

**Participants:** Umut Acar, Vitaly Aksenov, Arthur Charguéraud, Adrien Guatto [Université Paris Diderot], Mike Rainey, Filip Sieczkowski [University of Wrocław].

This year, the DeepSea team continued their work on granularity control techniques for parallel programs.

A first line of research is based on the use of programmer-supplied asymptotic complexity functions, combined with runtime measurements. This work first appeared at PPoPP 2018 [16] in the form of a brief announcement, and was subsequently accepted for publication at PPoPP 2019 as a full paper.

A second line of research, known as *heartbeat scheduling*, is based on instrumenting the runtime system so that parallel function calls are initially executed as normal function calls, by pushing a frame on the stack, and subsequently can be promoted and become independent threads. This research has been presented at PLDI 2018 [14].

### 7.3.6. Theory and analysis of concurrent algorithms

**Participant:** Vitaly Aksenov.

Vitaly Aksenov, in collaboration with Petr Kuznetsov (Télécom ParisTech) and Anatoly Shalyto (ITMO University), proved that no wait-free linearizable implementation of a stack using read, write, compare & swap and fetch & add operations can be help-free. This proof corrects a mistake in an earlier proof by Censor-Hillel et al. The result was published at the the International Conference on Networked Systems (NETYS 2018) [17].

Vitaly Aksenov, in collaboration with Dan Alistarh (IST Austria) and Petr Kuznetsov (Télécom ParisTech), worked on performance prediction for coarse-grained locking. They describe a simple model that can be used to predict the throughput of coarse-grained lock-based algorithms. They show that their model works well for CLH locks, and thus can be expected to work for other popular lock designs such as TTAS or MCS. This work appeared as a brief announcement at PODC 2018 [16].

The aforementioned results by Vitaly Aksenov are also covered in his Ph.D. manuscript [11].

## 7.4. The OCaml language and system

### 7.4.1. The OCaml system

**Participants:** Damien Doligez, Armaël Guéneau, Xavier Leroy, Luc Maranget, David Allsop [University of Cambridge], Florian Angeletti, Frédéric Bour [Facebook], Stephen Dolan [University of Cambridge], Alain Frisch [Lexifi], Jacques Garrigue [University of Nagoya], Sébastien Hinderer, Nicolás Ojeda Bär [Lexifi], Thomas Refis [Jane Street], Gabriel Scherer [team Parsifal], Mark Shinwell [Jane Street], Leo White [Jane Street], Jeremy Yallop [University of Cambridge].

This year, we released three versions of the OCaml system: versions 4.06.1 and 4.07.1 are minor releases that fix 7 and 8 issues, respectively; version 4.07.0 is a major release that introduces many improvements in usability and performance, and fixes about 40 issues. The main novelties are:

- The standard library modules were reorganized to appear as sub-modules of a new `Stdlib` module. The purpose of this reorganization is to facilitate the addition of new standard library modules while minimize risks of conflicts with user modules of the same name.
- Modules `Float` (floating-point operations) and `Seq` (sequences) were added to the standard library, taking advantage of the new organization mentioned above.
- Since 4.01, it has been possible to select a variant constructor or record field from a sub-module that is not opened in the current scope, if type information is available at the point of use. This now also works for GADT constructors.
- The GC now handles the accumulation of custom blocks in the minor heap better. This solves some memory-usage issues observed in code which allocates a large amount of small custom blocks, typically small bigarrays.

### *7.4.2. Package management infrastructure*

**Participant:** Damien Doligez.

This year, Damien Doligez has worked on the `opamcheck` tool, which is designed to check the compatibility of different versions of OCaml on the whole code base of `opam`, OCaml's package manager. As a by-product of this work, he has proposed numerous fixes to the `opam` package repository and to its dependency graph.

### *7.4.3. Work on the compiler's test suite and build system*

**Participant:** Sébastien Hinderer.

In 2018, Sébastien Hinderer has worked on the OCaml compiler's test suite. More precisely, he has finished porting over 800 tests in the compiler's test suite so that they can be run by the tool `ocamltest`, developed by Sébastien earlier. To achieve this, it has been necessary to extend both `ocamltest` and the domain-specific language that is used to describe how tests should be executed.

In addition, Sébastien has fixed and properly documented the procedure that is used to bootstrap the OCaml compiler. Being able to compile the compiler using itself is an important feature: it is crucial, for instance, when the compiler is released. In addition to fixing the bootstrap procedure, Sébastien has introduced a way to test this procedure through continuous integration, which guarantees that it will not be broken again in the future.

Finally, Sébastien has continued to improve and refactor the compiler's build system, and, most importantly, has replaced the hand-written configuration script by an `autoconf`-generated one, which will be part of the upcoming 4.08 release of OCaml. This represents an important step towards the ability to produce cross-compilers for OCaml, which has been a long-standing issue for the whole OCaml community.

### *7.4.4. Optimizing OCaml for satisfiability problems*

**Participants:** Sylvain Conchon [LRI, Univ. Paris-Saclay], Albin Coquereau [ENSTA-ParisTech], Mohamed Iguernlala [OCamlPro], Fabrice Le fessant [OCamlPro], Michel Mauny.

This work aims at improving the performance of the Alt-Ergo SMT solver, which is implemented in OCaml. For safety reasons, and to ease reasoning about its algorithms, the implementation of Alt-Ergo uses a functional programming style and persistent data structures, which are sometimes less efficient than imperative style and mutable data. Moreover, some efficient algorithms, such as CDCL SAT solvers, are naturally expressed in an imperative style.

Following our previous work on optimizing Alt-Ergo's built-in SAT solver, some efforts were needed to enable the comparison of our solver with other SMT solvers. We developed an OCaml library for parsing and type-checking SMT-LIB2. Since Alt-Ergo natively uses a polymorphic typing discipline, and since the community needs such advanced features, we proposed an extension of the SMT-LIB2 syntax where functions may be polymorphic.

The resulting new version of Alt-Ergo was presented at the 2018 SMT Workshop in Oxford [33]. Comparisons of Alt-Ergo with other SMT solvers, mainly developed in C++, took place during the competition that is associated with the workshop. They showed that Alt-Ergo's performance is similar to that of its competitors.

Albin Coquereau's Ph.D. defense is planned for Spring 2019.

### *7.4.5. Improvements in Menhir*

**Participant:** François Pottier.

In 2018, the OCaml parser of the OCaml compiler was migrated from `ocamlyacc` to Menhir, at last. François Pottier took this opportunity to partially clean up the parser, reducing redundancy by taking advantage of Menhir's features. In the future, we hope to continue to work on the OCaml parser by improving the quality of its syntax error messages.

This cleanup work was also an occasion to revisit Menhir's grammar description language: François Pottier designed and implemented a new input syntax for Menhir, which seems slightly more powerful and elegant than the previous syntax.

## 7.5. Software specification and verification

### 7.5.1. Formal reasoning about asymptotic complexity

**Participants:** Armaël Guéneau, Arthur Charguéraud [team Camus], François Pottier.

For a couple years, Armaël Guéneau, Arthur Charguéraud, François Pottier have been investigating the use of Separation Logic, extended with Time Credits, as an approach to the formal verification of the time complexity of OCaml programs. In particular, Armaël has developed in Coq a theory and a set of tactics that allow working with asymptotic complexity bounds. He has presented the main aspects of this work at the conference ESOP 2018 [21]. Furthermore, a key part of the machinery for working with asymptotic complexity bounds has been released as a standalone, reusable Coq library, procrastination. Armaël presented this library at the Coq Workshop in July 2018 [29].

In 2018, Armaël has worked on a more ambitious case study, namely a recent incremental cycle detection algorithm, whose amortized complexity analysis is nontrivial. A machine-checked proof has been completed; a paper is in preparation.

### 7.5.2. Time Credits and Time Receipts in Iris

**Participants:** Glen Mével, Jacques-Henri Jourdan [CNRS], François Pottier.

From March to August 2018, Glen Mével did an M2 internship at Gallium, where he was co-advised by Jacques-Henri Jourdan (CNRS) and François Pottier. Glen extended the program logic Iris with time credits and time receipts.

Time credits are a well-understood concept, and have been used in several papers already by Armaël Guéneau, Arthur Charguéraud, and François Pottier. However, because Iris is implemented and proved sound inside Coq, extending Iris with time credits requires a nontrivial proof, which Glen carried out, based on a program transformation which inserts "tick" instructions into the code. As an application of time credits, Glen verified inside Iris the correctness of Okasaki's notion of "debits", which allows reasoning about the time complexity of programs that use thunks.

Time receipts are a new concept, which (we showed) allows proving that certain undesirable events, such as integer overflows, cannot occur until a very long time has elapsed. Glen extended Iris with time receipts and proved the soundness of this extension. As an application of time credits and receipts together, Jacques-Henri Jourdan updated Charguéraud and Pottier's earlier verification of the Union-Find data structure [3] and proved that integer ranks cannot realistically overflow, even if they are stored using only $\log W$ bits, where $W$ is the number of bits in a machine word.

This work has been first submitted to POPL 2019, then (after significant revision) re-submitted to ESOP 2019.

### 7.5.3. Verified Interval Maps

**Participant:** François Pottier.

In the setting of ANR project Vocal, which aims to build a library of verified data structures for OCaml, François Pottier carried out a formal reconstruction of "interval maps". An interval map, a data structure proposed by Bonichon and Cuoq in 2010, represents a set of possible heaps, that is, a set of mappings of integer addresses to abstract values. Interval maps are used in the Frama-C program analysis tool. François Pottier re-implemented this data structure in Coq and carried out a formal verification of its main operations. This work, which represents about 4 months of work, remains unpublished at this time. It would be desirable to publish it and to envision its integration in Frama-C; this however requires further effort.

### *7.5.4. Chunked Sequences*

**Participants:** Émilie Guermeur, Arthur Charguéraud, François Pottier.

In June and July 2018, Émilie Guermeur, an undergraduate student at Carnegie Mellon University (Pittsburgh, USA) did a 6-week internship, co-advised by Arthur Charguéraud and François Pottier. She wrote a full-fledged OCaml implementation of "chunked sequences", a data structure which offers an efficient representation of sequences of elements. This data structure exists in two forms, a persistent form and an ephemeral (mutable) form; efficient conversion operations are offered. François Pottier subsequently implemented a test harness, based on afl-fuzz, which allowed us to submit Émilie's code to intensive testing and detect and fix a few bugs. This work is not yet published; we intend to pursue it in 2019, to publish the library and perhaps to verify it.

### *7.5.5. TLA+*

**Participants:** Damien Doligez, Leslie Lamport [Microsoft Research], Ioannis Filippidis, Martin Riener [team VeriDis], Stephan Merz [team VeriDis].

Damien Doligez is head of the "Tools for Proofs" team in the Microsoft-Inria Joint Centre. The aim of this project is to extend the TLA+ language with a formal language for hierarchical proofs, formalizing Lamport's ideas [36]. This requires building tools to help write TLA+ specifications and mechanically check proofs.

Since October 2018, Ioannis Filippidis has been working on extending the TLAPS tool to deal with proofs of temporal properties. Under some well-defined circumstances, an occurrence of the ENABLED operator applied to a formula $f$ can be replaced by a version of $f$ where the primed variables are replaced by new existentially-quantified variables. The result is a first-order formula that can be sent to one of TLAPS's first-order backends. This rewriting of ENABLED suffices to prove a large class of liveness properties. Ioannis has started implementing this in TLAPS.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

### *8.1.1. The Caml Consortium*

**Participants:** Damien Doligez [ **contact** ], Xavier Leroy, Michel Mauny, Didier Rémy.

The Caml Consortium is a formal structure where industrial and academic users of OCaml can support the development of the language and associated tools, express their specific needs, and contribute to the long-term stability of OCaml. Membership fees are used to fund specific developments targeted towards industrial users. Members of the Consortium automatically benefit from very liberal licensing conditions on the OCaml system, allowing for instance the OCaml compiler to be embedded within proprietary applications.

The Consortium currently has 15 member companies:

- Aesthetic Integration
- Ahrefs
- Be Sport
- Bloomberg
- CEA
- Citrix
- Docker
- Esterel Technologies
- Facebook
- Jane Street
- Kernelyze LLC
- LexiFi
- Microsoft
- OCamlPro
- SimCorp

For a complete description of this structure, please refer to https://ocaml.org/consortium/index.html.

The Caml Consortium is being gradually phased out. In the future, it should be entirely replaced by the OCaml Foundation, described next (§8.1.2).

### *8.1.2. The OCaml Foundation*

**Participant:** Michel Mauny.

In June 2018, Michel Mauny created the OCaml Software Foundation (OCSF), a structure sheltered by the Inria Foundation. The OCSF now has a few patrons. With the help of Yann Régis-Gianas, it is running the Learn-OCaml project, which aims at developing the usage of OCaml in higher education. A paper that presents the project has been accepted for publication at JFLA 2019 [20]. The OCaml Software Foundation and the Learn-OCaml project have been presented at the 2018 OCaml workshop.

The OCaml Software Foundation is expecting more patrons at the beginning of 2019, and shall organize meetings where donors discuss and produce suggestions for actions of general interest to be funded.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR projects*

#### *9.1.1.1. Vocal*

**Participants:** Armaël Guéneau, Xavier Leroy, François Pottier.

The "Vocal" project (2015–2020) aims at developing the first mechanically verified library of efficient general-purpose data structures and algorithms. It is funded by *Agence Nationale de la Recherche* under its "appel à projets générique 2015".

A first release of the library has been published in December 2018. It contains a small number of verified data structures, including resizable vectors, hash tables, priority queues, and Union-Find.

### *9.1.2. FUI Projects*

#### *9.1.2.1. Secur-OCaml*

**Participants:** Damien Doligez, Fabrice Le Fessant.

The "Secur-OCaml" project (2015–2018) has been coordinated by the OCamlPro company, with a consortium focusing on the use of OCaml in security-critical contexts, while OCaml is currently mostly used in safety-critical contexts. Gallium has been involved in this project to integrate security features in the OCaml language, to build a new independent interpreter for the language, and to update the recommendations for developers issued by the former LaFoSec project of ANSSI. The end-of-project meeting took place in September 2018.

## 9.2. European Initiatives

### *9.2.1. FP7 & H2020 Projects*

#### *9.2.1.1. Deepsea*

**Participants:** Umut Acar, Vitaly Aksenov, Arthur Charguéraud, Adrien Guatto, Michael Rainey.

The Deepsea project (2013–2018) is coordinated by Umut Acar and funded by FP7 as an ERC Starting Grant. Its objective is to develop abstractions, algorithms and languages for parallelism and dynamic parallelism, with applications to problems on large data sets.

### *9.2.2. ITEA3 Projects*

*9.2.2.1. Assume*

**Participants:** Gergö Barany, Xavier Leroy, Luc Maranget.

ASSUME (2015–2018) is an ITEA3 project involving France, Germany, Netherlands, Turkey and Sweden. The French participants are coordinated by Jean Souyris (Airbus) and include Airbus, Kalray, Sagem, ENS Paris, and Inria Paris. The goal of the project is to investigate the usability of multicore and manycore processors for critical embedded systems. Our involvement in this project focuses on the formalisation and verification of memory models and of automatic code generators from reactive languages, as well as on extensions to the CompCert C compiler.

## 9.3. International Initiatives

### *9.3.1. Informal International Partners*

- Princeton University: interactions between the CompCert verified C compiler and the Verified Software Toolchain developed at Princeton.
- The University of Cambridge and ARM Ltd, Cambridge and Imperial College London: formal modeling and testing of weak memory models.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Selection*

*10.1.1.1. Member of the Conference Program Committees*

Xavier Leroy was on the program committee of CADO 2018, the special session on Compiler, Architecture, Design and Optimization of the 16th International Conference on High Performance Computing and Simulation.

Michel Mauny has been a member of the program committee of the International Symposium on Image, Video and Communications (ISIVC 2018).

François Pottier was a member of the program committee of ICFP 2018, the ACM International Conference on Functional Programming.

Didier Rémy was a member of the program committee of FLOPS 2018, the 14th International Symposium on Functional and Logic Programming.

### *10.1.2. Journal*

*10.1.2.1. Member of the Editorial Boards*

Xavier Leroy is area editor (programming languages) for Journal of the ACM. He is a member of the editorial board of Journal of Automated Reasoning.

Until September 2018, Michel Mauny has been a member of the steering committee of the OCaml workshop.

François Pottier is a member of the ICFP steering committee and a member of the editorial boards of the Journal of Functional Programming and the Proceedings of the ACM on Programming Languages.

Didier Rémy is a member of the steering committee of the ML Family workshop.

### *10.1.3. Research Administration*

In 2018, Michel Mauny was chairman of the Scientific Committee of the Caml Consortium. He organized its annual meeting in December 2018.

Since May 2018, Michel Mauny has been Chief Executive Officer of the Inria Foundation.

François Pottier is a member of Inria Paris' *Commission de Développement Technologique* and the president of Inria Paris' *Comité de Suivi Doctoral*.

Didier Rémy is *Deputy Scientific Director* (ADS) in charge of *Algorithmics, Programming, Software and Architecture*.

Didier Rémy is Inria's delegate in the pedagogical board of the *Master Parisien de Recherche en Informatique* (MPRI).

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master (M2): "Proofs of Programs", Jean-Marie Madiot, 18 HETD, Université Paris Diderot, France.

Master (M2): "Semantics, languages and algorithms for multi-core programming", Luc Maranget, 18 HETD, Université Paris Diderot, France.

Master (M2): "Functional programming and type systems", François Pottier, 18 HETD, Université Paris Diderot, France.

Master (M2): "Functional programming and type systems", Didier Rémy, 18 HETD, Université Paris Diderot, France.

Licence (L3): Jean-Marie Madiot, "Les principes des langages de programmation", 40 HETD, École Polytechnique, France.

Master (M1): Michel Mauny, "Principles of Programming Languages", 32 HETD, ENSTA-ParisTech, France.

Open lectures: Xavier Leroy, *Programmer = démontrer? La correspondance de Curry-Howard aujourd'hui*, 16 HETD, Collège de France, France.

### *10.2.2. Supervision*

PhD: Vitaly Aksenov, "Synchronization Costs in Parallel Programs and Concurrent Data Structures", ITMO University of Saint Petersburg (Russia) and Université Paris Diderot, September 26, 2018, advised by Petr Kuznetsov and Anatoly Shalyto [11].

PhD: Pierrick Couderc, "Vérification des résultats de l'inférence du compilateur OCaml", Université Paris-Saclay, October 23, 2018, advised by Michel Mauny et Fabrice Le Fessant [34].

PhD in progress: Albin Coquereau, "Amélioration de performances pour le solveur SMT Alt-Ergo: conception d'outils d'analyse, optimisations et structures de données efficaces pour OCaml," Université Paris-Saclay, since October 2015, advised by Michel Mauny, Sylvain Conchon (LRI, Université Paris-Sud) and Fabrice Le Fessant.

PhD in progress: Armaël Guéneau, "Towards Machine-Checked Time Complexity Analyses", Université Paris Diderot, since September 2016, advised by Arthur Charguéraud and François Pottier.

PhD in progress: Glen Mével, "Towards a system for proving the correctness of concurrent Multicore OCaml programs", Université Paris Diderot, since November 2018, advised by Jacques-Henri Jourdan and François Pottier.

PhD in progress: Naomi Testard, "Reasoning about Effect Handlers and Cooperative Concurrency", Université Paris Diderot, since January 2017, advised by François Pottier.

PhD in progress: Thomas Williams, "Putting Ornaments into practice", Université Paris Diderot, since September 2014, advised by Didier Rémy.

### *10.2.3. Juries*

Xavier Leroy was a member of the jury for the Habilitation defense of Julien Signoles (Université Paris Sud, July 2018).

Xavier Leroy chaired the jury for the Ph.D. defense of Mario Pereira (Université Paris Sud, December 2018).

François Pottier was a reviewer for Steven Keuchel's PhD thesis (Ghent University), defended on June 5, 2018.

François Pottier was a reviewer for Martin Clochard's PhD thesis (Université Paris-Saclay), defended on March 30, 2018.

## 10.3. Popularization

### *10.3.1. Articles and contents*

- For online publications (Interstices*, Images des Maths, Binaire, Wikipedia), and more widely blog articles

  Xavier Leroy wrote a short introduction to software sciences in general and to his lectures at Collège de France. This text was published by the "Binaire" blog of *Le Monde* [30].

### *10.3.2. Interventions*

Gergö Barany gave a talk titled "Finding Missed Optimizations in LLVM (and other compilers)" at the 2018 European LLVM Developers Meeting, explaining his research on testing the quality of compiler optimizations to practitioners in compiler development.

# 11. Bibliography

## Major publications by the team in recent years

[1] J. ALGLAVE, L. MARANGET, M. TAUTSCHNIG.*Herding cats: modelling, simulation, testing, and data-mining for weak memory*, in "ACM Transactions on Programming Languages and Systems", 2014, vol. 36, n° 2, article no 7, http://dx.doi.org/10.1145/2627752

[2] T. BALABONSKI, F. POTTIER, J. PROTZENKO.*The design and formalization of Mezzo, a permission-based programming language*, in "ACM Transactions on Programming Languages and Systems", 2016, vol. 38, n° 4, p. 14:1–14:94, http://doi.acm.org/10.1145/2837022

[3] A. CHARGUÉRAUD, F. POTTIER.*Verifying the Correctness and Amortized Complexity of a Union-Find Implementation in Separation Logic with Time Credits*, in "Journal of Automated Reasoning", September 2017 [*DOI :* 10.1007/s10817-017-9431-7], https://hal.inria.fr/hal-01652785

[4] K. CHAUDHURI, D. DOLIGEZ, L. LAMPORT, S. MERZ.*Verifying Safety Properties With the TLA+ Proof System*, in "Automated Reasoning, 5th International Joint Conference, IJCAR 2010", Lecture Notes in Computer Science, Springer, 2010, vol. 6173, p. 142–148, http://dx.doi.org/10.1007/978-3-642-14203-1_12

[5] J. CRETIN, D. RÉMY.*System F with Coercion Constraints*, in "CSL-LICS 2014: Computer Science Logic / Logic In Computer Science", ACM, 2014, article no 34, http://dx.doi.org/10.1145/2603088.2603128

[6] J.-H. JOURDAN, V. LAPORTE, S. BLAZY, X. LEROY, D. PICHARDIE.*A Formally-Verified C Static Analyzer*, in "POPL'15: 42nd ACM Symposium on Principles of Programming Languages", ACM Press, January 2015, p. 247-259, http://dx.doi.org/10.1145/2676726.2676966

[7] D. LE BOTLAN, D. RÉMY.*Recasting MLF*, in "Information and Computation", 2009, vol. 207, n° 6, p. 726–785, http://dx.doi.org/10.1016/j.ic.2008.12.006

[8] X. LEROY.*A formally verified compiler back-end*, in "Journal of Automated Reasoning", 2009, vol. 43, n° 4, p. 363–446, http://dx.doi.org/10.1007/s10817-009-9155-4

[9] X. LEROY.*Formal verification of a realistic compiler*, in "Communications of the ACM", 2009, vol. 52, n° 7, p. 107–115, http://doi.acm.org/10.1145/1538788.1538814

[10] N. POUILLARD, F. POTTIER.*A unified treatment of syntax with binders*, in "Journal of Functional Programming", 2012, vol. 22, n° 4–5, p. 614–704, http://dx.doi.org/10.1017/S0956796812000251

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[11] V. AKSENOV.*Synchronization Costs in Parallel Programs and Concurrent Data Structures*, ITMO University ; Paris Diderot University, September 2018, https://hal.inria.fr/tel-01887505

## Articles in International Peer-Reviewed Journal

[12] T. WILLIAMS, D. RÉMY.*A Principled Approach to Ornamentation in ML*, in "Proceedings of the ACM on Programming Languages", January 2018, p. 1-30 [*DOI :* 10.1145/3158109], https://hal.inria.fr/hal-01666104

## International Conferences with Proceedings

[13] U. A. ACAR, V. AKSENOV, A. CHARGUÉRAUD, M. RAINEY.*Provably and Practically Efficient Granularity Control*, in "PPoPP 2019 - Principles and Practice of Parallel Programming", Washington DC, United States, February 2019 [*DOI :* 10.1145/3293883.3295725], https://hal.inria.fr/hal-01973285

[14] U. A. ACAR, A. CHARGUÉRAUD, A. GUATTO, M. RAINEY, F. SIECZKOWSKI.*Heartbeat scheduling: provable efficiency for nested parallelism*, in "PLDI'18 - 39th ACM SIGPLAN Conference on Programming Language Design and Implementation", Philadelphia, United States, ACM Press, June 2018 [*DOI :* 10.1145/3192366.3192391], https://hal.inria.fr/hal-01937946

[15] V. AKSENOV, U. A. ACAR, A. CHARGUÉRAUD, M. RAINEY.*Poster: Performance challenges in modular parallel programs*, in "PPoPP 2018 - 23rd ACM SIGPLAN Annual Symposium on Principles and Practice of Parallel Programming", Vienna, Austria, February 2018, vol. 18 [*DOI :* 10.1145/3178487.3178516], https://hal.inria.fr/hal-01887717

[16] V. AKSENOV, D. ALISTARH, P. KUZNETSOV.*Brief Announcement: Performance Prediction for Coarse-Grained Locking*, in "PODC 2018 - ACM Symposium on Principles of Distributed Computing", Egham, United Kingdom, July 2018 [*DOI :* 10.1145/3212734.3212785], https://hal.inria.fr/hal-01887733

[17] V. AKSENOV, P. KUZNETSOV, A. SHALYTO.*On Helping and Stacks*, in "The International Conference on Networked Systems", Essaouira, Morocco, May 2018, https://hal.inria.fr/hal-01888607

[18] J. ALGLAVE, L. MARANGET, P. MCKENNEY, A. PARRI, A. STERN.*Frightening Small Children and Disconcerting Grown-ups: Concurrency in the Linux Kernel*, in "ASPLOS2018 - 23rd ACM International Conference on Architectural Support for Programming Languages and Operating Systems", Williamsburg, VA, United States, March 2018 [*DOI :* 10.1145/3173162.3177156], https://hal.inria.fr/hal-01873636

[19] G. BARANY.*Finding Missed Compiler Optimizations by Differential Testing*, in "CC'18 - 27th International Conference on Compiler Construction", Vienna, Austria, February 2018 [*DOI :* 10.1145/3178372.3179521], https://hal.inria.fr/hal-01682683

[20] C. BOZMAN, B. CANOU, R. DI COSMO, P. COUDERC, L. GESBERT, G. HENRY, F. LE FESSANT, M. MAUNY, C. MOREL, L. PEYROT.*Learn-OCaml : un assistant à l'enseignement d'OCaml*, in "Journées Francophones des Langages Applicatifs (JFLA)", Les Rousses, France, January 2019, https://hal.inria.fr/hal-01962838

[21] A. GUÉNEAU, A. CHARGUÉRAUD, F. POTTIER.*A Fistful of Dollars: Formalizing Asymptotic Complexity Claims via Deductive Program Verification*, in "ESOP 2018 - 27th European Symposium on Programming", Thessaloniki, Greece, A. AHMED (editor), LNCS - Lecture Notes in Computer Science, Springer, April 2018, vol. 10801, Held as Part of the European Joint Conferences on Theory and Practice of Software, ETAPS 2018 [*DOI :* 10.1007/978-3-319-89884-1_19], https://hal.inria.fr/hal-01926485

[22] D. KÄSTNER, J. BARRHO, U. WÜNSCHE, M. SCHLICKLING, B. SCHOMMER, M. SCHMIDT, C. FERDINAND, X. LEROY, S. BLAZY.*CompCert: Practical Experience on Integrating and Qualifying a Formally Verified Optimizing Compiler*, in "ERTS2 2018 - 9th European Congress Embedded Real-Time Software and Systems", Toulouse, France, 3AF, SEE, SIE, January 2018, p. 1-9, https://hal.inria.fr/hal-01643290

[23] B. SCHOMMER, C. CULLMANN, G. GEBHARD, X. LEROY, M. SCHMIDT, S. WEGENER.*Embedded Program Annotations for WCET Analysis*, in "WCET 2018: 18th International Workshop on Worst-Case Execution Time Analysis", Barcelona, Spain, Dagstuhl Publishing, July 2018, vol. 63 [*DOI :* 10.4230/OASIcs.WCET.2018.8], https://hal.inria.fr/hal-01848686

### National Conferences with Proceeding

[24] G. BARANY, G. SCHERER.*Génération aléatoire de programmes guidée par la vivacité*, in "JFLA 2018 - Journées Francophones des Langages Applicatifs", Banyuls-sur-Mer, France, January 2018, https://hal.inria.fr/hal-01682691

### Conferences without Proceedings

[25] G. BARANY.*A more precise, more correct stack and register model for CompCert*, in "LOLA 2018 - Syntax and Semantics of Low-Level Languages 2018", Oxford, United Kingdom, July 2018, https://hal.inria.fr/hal-01799629

### Research Reports

[26] X. LEROY, D. DOLIGEZ, A. FRISCH, J. GARRIGUE, D. RÉMY, J. VOUILLON.*The OCaml system release 4.07: Documentation and user's manual*, Inria, July 2018, p. 1-752, https://hal.inria.fr/hal-00930213

[27] X. LEROY.*The CompCert C verified compiler: Documentation and user's manual: Version 3.4*, Inria, September 2018, p. 1-77, https://hal.inria.fr/hal-01091802

[28] F. PESSAUX, D. DOLIGEZ.*Compiling Programs and Proofs: FoCaLiZe Internals*, Ensta ParisTech, May 2018, https://hal.archives-ouvertes.fr/hal-01801276

### Other Publications

[29] A. GUÉNEAU.*Procrastination: A proof engineering technique*, July 2018, Coq Workshop 2018, The Coq Workshop 2018 is a part of FLoC 2018, https://hal.inria.fr/hal-01962659

[30] X. LEROY.*À la recherche du logiciel parfait*, November 2018, Post on the "Binaire" popular science blog of Le Monde, https://hal.inria.fr/hal-01966252

## References in notes

[31] C. BASTOUL.*Code Generation in the Polyhedral Model Is Easier Than You Think*, in "PACT'04: Proceedings of the 13th International Conference on Parallel Architectures and Compilation Techniques", IEEE Computer Society, 2004, p. 7–16

[32] V. BENZAKEN, G. CASTAGNA, A. FRISCH.*CDuce: an XML-centric general-purpose language*, in "Proceedings of the Eighth ACM SIGPLAN International Conference on Functional Programming", C. RUNCIMAN, O. SHIVERS (editors), ACM, 2003, p. 51–63, https://www.lri.fr/~benzaken/papers/icfp03.ps

[33] S. CONCHON, A. COQUEREAU, M. IGUERNELALA, A. MEBSOUT.*Alt-Ergo 2.2*, in "Proceedings of the 16th International Workshop on Satisfiability Modulo Theories, SMT 2018", Oxford, UK, 2018, https://github.com/OCamlPro/alt-ergo/blob/next/publications/Alt-Ergo-2.2–SMT-Workshop-2018.pdf

[34] P. COUDERC.*Vérification des résultats de l'inférence de types du langage OCaml*, Université Paris-Saclay, 2018

[35] H. HOSOYA, B. C. PIERCE.*XDuce: A Statically Typed XML Processing Language*, in "ACM Transactions on Internet Technology", 2003, vol. 3, n$^{o}$ 2, p. 117–148, http://doi.acm.org/10.1145/767193.767195

[36] L. LAMPORT.*How to write a 21st century proof*, in "Journal of Fixed Point Theory and Applications", 2012, vol. 11, p. 43–63, http://dx.doi.org/10.1007/s11784-012-0071-6

[37] X. LEROY, D. DOLIGEZ, A. FRISCH, J. GARRIGUE, D. RÉMY, J. VOUILLON.*The Objective Caml system, documentation and user's manual – release 4.07*, Inria, July 2018, http://caml.inria.fr/pub/docs/manual-ocaml-4.07/

[38] X. LEROY.*Java bytecode verification: algorithms and formalizations*, in "Journal of Automated Reasoning", 2003, vol. 30, n$^{o}$ 3–4, p. 235–269, http://dx.doi.org/10.1023/A:1025055424017

[39] B. C. PIERCE.*Types and Programming Languages*, MIT Press, 2002

[40] F. POTTIER.*Simplifying subtyping constraints: a theory*, in "Information and Computation", 2001, vol. 170, n$^{o}$ 2, p. 153–183, http://gallium.inria.fr/~fpottier/publis/fpottier-ic01.ps.gz

[41] F. POTTIER, V. SIMONET.*Information Flow Inference for ML*, in "ACM Transactions on Programming Languages and Systems", January 2003, vol. 25, n$^{o}$ 1, p. 117–158, http://dx.doi.org/10.1145/596980.596983

[42] D. RÉMY, J. VOUILLON.*Objective ML: A simple object-oriented extension to ML*, in "24th ACM Conference on Principles of Programming Languages", ACM Press, 1997, p. 40–53, http://gallium.inria.fr/~remy/ftp/objective-ml!popl97.pdf

[43] T. WILLIAMS, D. RÉMY.*A Principled Approach to Ornamentation in ML*, Inria, November 2017, n$^{o}$ RR-9117 [*DOI :* 10.1145/NNNNNNN.NNNNNNN], https://hal.inria.fr/hal-01628060

# Project-Team GANG

## Networks, Graphs and Algorithms

IN COLLABORATION WITH: Institut de Recherche en Informatique Fondamentale

IN PARTNERSHIP WITH:
**CNRS**

**Université Denis Diderot (Paris 7)**

RESEARCH CENTER
**Paris**

THEME
**Networks and Telecommunications**

# Table of contents

# Project-Team GANG

*Creation of the Project-Team: 2007 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.2. - Networks
A1.2.3. - Routing
A1.2.9. - Social Networks
A1.3. - Distributed Systems
A3.5. - Social networks
A3.5.1. - Analysis of large graphs
A6.1.3. - Discrete Modeling (multi-agent, people centered)
A7.1. - Algorithms
A7.1.3. - Graph algorithms
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A8.7. - Graph theory
A8.8. - Network science

**Other Research Topics and Application Domains:**

B1.1.6. - Evolutionnary biology
B1.1.10. - Systems and synthetic biology
B6.3.2. - Network protocols
B6.3.4. - Social Networks
B7.2. - Smart travel

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Laurent Viennot [Team leader, Inria, Senior Researcher, HDR]
Pierre Fraigniaud [CNRS, Senior Researcher, HDR]
Amos Korman [CNRS, Senior Researcher, HDR]
Adrian Kosowski [Inria, Researcher, HDR]

**Faculty Members**

Michel Habib [Team leader, Univ Denis Diderot, Professor emeritus, HDR]
Yacine Boufkhad [Univ Denis Diderot, Associate Professor]
Pierre Charbit [Univ Denis Diderot, Associate Professor]
Fabien de Montgolfier [Univ Denis Diderot, Associate Professor]
Carole Delporte-Gallet [Univ Denis Diderot, Professor, HDR]
Hugues Fauconnier [Univ Denis Diderot, Professor, HDR]

**PhD Students**

Simon Collet [ERC BDA]
Lucas Boczkowski [ERC BDA]
Laurent Feuilloley [ENS Paris-Saclay]
Mengchuan Zou [Inria]

**Administrative Assistant**
Christine Anocq [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

GANG focuses on algorithm design for large scale networks using structural properties of these networks. Application domains include the development of optimized protocols for large dynamic networks such as mobile networks or overlay networks over Internet. This includes for instance peer-to-peer applications, or the navigability of social networks. GANG tools come from recent advances in the field of graph algorithms, both in centralized and distributed settings. In particular, this includes graph decomposition and geometric properties (such as low doubling dimension, low dimension embedding, etc.). Today, the management of large networks, Internet being the reference, is best effort. However, the demand for mobility (ad hoc networks, wireless connectivity, etc.) and for dynamicity (node churn, fault tolerance, etc.) is increasing. In this distributed setting, it becomes necessary to design a new generation of algorithms and protocols to face the challenge of large scale mobility and dynamicity. In the mean time, recent and sophisticated theoretical results have emerged, offering interesting new tracks for managing large networks. These results concern centralized and decentralized algorithms for solving key problems in communication networks, including routing, but also information retrieval, localization, or load balancing. They are mainly based on structural properties observed in most of real networks: approximate topology with low dimension metric spaces, low treewidth, low doubling dimension, graph minor freeness, etc. In addition, graph decomposition techniques have recently progressed. The scientific community has now tools for optimizing network management. First striking results include designing overlay networks for peer-to-peer systems and understanding the navigability of large social networks.

# 3. Research Program

## 3.1. Graph and Combinatorial Algorithms

We focus on two approaches for designing algorithms for large graphs: decomposing the graph and relying on simple graph traversals.

### *3.1.1. Graph Decompositions*

We study new decompositions schemes such as 2-join, skew partitions and others partition problems. These graph decompositions appeared in the structural graph theory and are the basis of some well-known theorems such as the Perfect Graph Theorem. For these decompositions there is a lack of efficient algorithms. We aim at designing algorithms working in $O(nm)$ since we think that this could be a lower bound for these decompositions.

### *3.1.2. Graph Search*

We more deeply study multi-sweep graph searches. In this domain a graph search only yields a total ordering of the vertices which can be used by the subsequent graph searches. This technique can be used on huge graphs and do not need extra memory. We already have obtained preliminary results in this direction and many well-known graph algorithms can be put in this framework. The idea behind this approach is that each sweep discovers some structure of the graph. At the end of the process either we have found the underlying structure (for example an interval representation for an interval graph) or an approximation of it (for example in hard discrete optimization problems). We envision applications to exact computations of centers in huge graphs, to underlying combinatorial optimization problems, but also to networks arising in biology.

### *3.1.3. Graph Exploration*

In the course of graph exploration, a mobile agent is expected to regularly visit all the nodes of an unknown network, trying to discover all its nodes as quickly as possible. Our research focuses on the design and analysis of agent-based algorithms for exploration-type problems, which operate efficiently in a dynamic network environment, and satisfy imposed constraints on local computational resources, performance, and resilience. Our recent contributions in this area concern the design of fast deterministic algorithms for teams of agents operating in parallel in a graph, with limited or no persistent state information available at nodes. We plan further studies to better understand the impact of memory constraints and of the availability of true randomness on efficiency of the graph exploration process.

## 3.2. Distributed Computing

The distributed computing community can be viewed as a union of two sub-communities. This is also true in our team. Although they have interactions, they are disjoint enough not to leverage each others' results. At a high level, one is mostly interested in timing issues (clock drifts, link delays, crashes, etc.) while the other one is mostly interested in spatial issues (network structure, memory requirements, etc.). Indeed, one sub-community is mostly focusing on the combined impact of asynchronism and faults on distributed computation, while the other addresses the impact of network structural properties on distributed computation. Both communities address various forms of computational complexity, through the analysis of different concepts. This includes, e.g., failure detectors and wait-free hierarchy for the former community and compact labeling schemes, and computing with advice for the latter community. We have an ambitious project to achieve the reconciliation between the two communities by focusing on the same class of problems, the yes/no-problems, and establishing the scientific foundations for building up a consistent theory of computability and complexity for distributed computing. The main question addressed is therefore: is the absence of globally coherent computational complexity theories covering more than fragments of distributed computing, inherent to the field? One issue is obviously the types of problems located at the core of distributed computing. Tasks like consensus, leader election, and broadcasting are of very different nature. They are not *yes-no* problems, neither are they minimization problems. Coloring and Minimal Spanning Tree are optimization problems but we are often more interested in constructing an optimal solution than in verifying the correctness of a given solution. Still, it makes full sense to analyze the *yes-no* problems corresponding to checking the validity of the output of tasks. Another issue is the power of individual computation. The FLP impossibility result as well as Linial's lower bound hold independently from the individual computational power of the involved computing entities. For instance, the individual power of solving NP-hard problems in constant time would not help overcoming these limits, which are inherent to the fact that computation is distributed. A third issue is the abundance of models for distributed computing frameworks, from shared memory to message passing, spanning all kinds of specific network structures (complete graphs, unit-disk graphs, etc.) and/or timing constraints (from complete synchronism to full asynchronism). There are however models, typically the wait-free model and the LOCAL model, which, though they do not claim to reflect accurately real distributed computing systems, enable focusing on some core issues. Our research program is ongoing to carry many important notions of Distributed Computing into a *standard* computational complexity.

## 3.3. Network Algorithms and Analysis

Based on our scientific foundation on both graph algorithms and distributed algorithms, we plan to analyze the behavior of various networks such as future Internet, social networks, overlay networks resulting from distributed applications or online social networks.

### *3.3.1. Information Dissemination*

One of the key aspects of networks resides in the dissemination of information among the nodes. We aim at analyzing various procedures of information propagation from dedicated algorithms to simple distributed schemes such as flooding. We also consider various models, e.g. where noise can alter information as it propagates or where memory of nodes is limited.

### 3.3.2. Routing Paradigms

We try to explore new routing paradigms such as greedy routing in social networks for example. We are also interested in content centric networking where routing is based on content name rather than content address. One of our target is multiple path routing: how to design forwarding tables providing multiple disjoint paths to the destination?

### 3.3.3. Beyond Peer-to-Peer

Based on our past experience of peer-to-peer application design, we would like to broaden the spectrum of distributed applications where new efficient algorithms can be designed and their analysis can be performed. We especially target online social networks as we see them as collaborative tools for exchanging information. A basic question resides in making the right connections for gathering filtered and accurate information with sufficient coverage.

### 3.3.4. SAT and Forwarding Information Verification

As forwarding tables of networks grow and are sometimes manually modified, the problem of verifying them becomes critical and has recently gained in interest. Some problems that arise in network verification such as loop detection for example, may be naturally encoded as Boolean Satisfiability problems. Beside theoretical interest in complexity proofs, this encoding allows one to solve these problems by taking advantage of the many efficient Satisfiability testing solvers. Indeed, SAT solvers have proved to be very efficient in solving problems coming from various areas (Circuit Verification, Dependency and Conflicts in Software distributions...) and encoded in Conjunctive Normal Form. To test an approach using SAT solvers in network verification, one needs to collect data sets from a real network and to develop good models for generating realistic networks. The technique of encoding and the solvers themselves need to be adapted to this kind of problems. All this represents a rich experimental field of future research.

### 3.3.5. Network Analysis

Finally, we are interested in analyzing the structural properties of practical networks. This can include diameter computation or ranking of nodes. As we mostly consider large networks, we are often interested in efficient heuristics. Ideally, we target heuristics that give exact answers and are reasonably fast in practice although fast computation time is not guaranteed for all networks. We have already designed such heuristics for diameter computation; understanding the structural properties that enable fast computation time in practice is still an open question.

# 4. Application Domains

## 4.1. Large scale networks

Application domains include evaluating Internet performances, the design of new peer-to-peer applications, enabling large scale networks, and developping tools for transportation networks.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

**WENDY: Workshop on Emergent Algorithms and Network Dynamics**

GANG/Inria Paris was the institutional organizer of WENDY workshop at Institut Henri-Poincaré, Paris, October 10-11, 2018, https://wendy.paris (chair: Adrian Kosowski).

The goal of the project was to facilitate the exchange of ideas between researchers working on distributed computing theory, modeling random structures, and discrete dynamical systems.

The main theme of the workshop was programming local interaction dynamics on networks, so as to obtain the desired emergent effects on the system as a whole. Central topics included:

- Evolving graph models and dynamics on random graphs
- Bio-inspired computing and computing with biological agents
- Chemical reaction networks
- Markovian and non-Markovian processes on networks.

**BDA: Workshop on Biological Distributed Algorithms**

Amos Korman chaired the organizing committee and co-chaired the program committee of the 6th workshop on Biological Distributed Algorithms (BDA, http://www.snl.salk.edu/~navlakha/BDA2018/), co-located with ACM PODC in London on July 23rd, 2018.

BDA was focused on the relationships between distributed computing and distributed biological systems and in particular, on analysis and case studies that combine the two. Such research can lead to better understanding of the behavior of the biological systems while at the same time developing novel algorithms that can be used to solve basic distributed computing problems.

The workshop featured 6 invited talks and over a dozen accepted contributed submissions, with generous financial support offered to participants by Amos Korman's ERC grant.

# 6. New Software and Platforms

## 6.1. big-graph-tools

KEYWORD: Graph algorithmics
FUNCTIONAL DESCRIPTION: Gang is developping a software for big graph manipulation. A preliminary library offering diameter and skeleton computation. This library was used to compute the diameters of the worldwide road network (200M edges) and the largest strongly connected component of the Twitter follower-followee graph (23G edges).

- Contact: Laurent Viennot
- URL: https://who.rocq.inria.fr/Laurent.Viennot/dev/big-graph-tools/

## 6.2. GRPH

*The high performance graph library for Java*
KEYWORDS: Graph - Graph algorithmics - Java
FUNCTIONAL DESCRIPTION: Grph is an open-source Java library for the manipulation of graphs. Its design objectives are to make it portable, simple to use/extend, computationally/memory efficient, and, according to its initial motivation: useful in the context of graph experimentation and network simulation. Grph also has the particularity to come with tools like an evolutionary computation engine, a bridge to linear programming solvers, a framework for distributed computing, etc.

Grph offers a very general model of graphs. Unlike other graph libraries which impose the user to first decide if he wants to deal with directed, undirected, hyper (or not) graphs, the model offered by Grph is unified in a general class that supports mixed graphs made of undirected and directed simple and hyper edges. Grph achieves great efficiency through the use of multiple code optimization techniques such as multi-core parallelism, caching, adequate data structures, use of primitive objects, exploitation of low-level processor caches, on-the-fly compilation of specific C/C++ code, etc. Grph attempts to access the Internet in order to check if a new version is available and to report who is using it (login name and hostname). This has no impact whatsoever on performance and security.

- Participants: Aurélien Lancin, David Coudert, Issam Tahiri, Luc Hogie and Nathann Cohen
- Contact: Luc Hogie
- URL: http://www.i3s.unice.fr/~hogie/grph/

# 7. New Results

## 7.1. Graph and Combinatorial Algorithms

### 7.1.1. Random Walks with Multiple Step Lengths

In nature, search processes that use randomly oriented steps of different lengths have been observed at both the microscopic and the macroscopic scales. Physicists have analyzed in depth two such processes on grid topologies: *Intermittent Search*, which uses two step lengths, and *Lévy Walk*, which uses many. Taking a computational perspective, in [26] we consider the number of distinct step lengths $k$ as a *complexity measure* of the considered process. Our goal is to understand what is the optimal achievable time needed to cover the whole terrain, for any given value of $k$. Attention is restricted to dimension one, since on higher dimensions, the simple random walk already displays a quasi linear cover time.

We say $X$ is a *$k$-intermittent search* on the one dimensional $n$-node cycle if there exists a probability distribution $\mathbf{p} = (p_i)_{i=1}^{k}$, and integers $L_1, L_2, ..., L_k$, such that on each step $X$ makes a jump $\pm L_i$ with probability $p_i$, where the direction of the jump ($+$ or $-$) is chosen independently with probability $1/2$. When performing a jump of length $L_i$, the process consumes time $L_i$, and is only considered to visit the last point reached by the jump (and not any other intermediate nodes). This assumption is consistent with biological evidence, in which entities do not search while moving ballistically. We provide upper and lower bounds for the cover time achievable by $k$-intermittent searches for any integer $k$. In particular, we prove that in order to reduce the cover time $\Theta(n^2)$ of a simple random walk to $\widetilde{\Theta}(n)$, roughly $\frac{\log n}{\log \log n}$ step lengths are both necessary and sufficient, and we provide an example where the lengths form an exponential sequence.

In addition, inspired by the notion of intermittent search, we introduce the *Walk or Probe* problem, which can be defined with respect to arbitrary graphs. Here, it is assumed that querying (probing) a node takes significantly more time than moving to a random neighbor. Hence, to efficiently probe all nodes, the goal is to balance the time spent walking randomly and the time spent probing. We provide preliminary results for connected graphs and regular graphs.

### 7.1.2. Searching a Tree with Permanently Noisy Advice

In [16], we consider a search problem on trees using unreliable guiding instructions. Specifically, an agent starts a search at the root of a tree aiming to find a treasure hidden at one of the nodes by an adversary. Each visited node holds information, called *advice*, regarding the most promising neighbor to continue the search. However, the memory holding this information may be unreliable. Modeling this scenario, we focus on a probabilistic setting. That is, the advice at a node is a pointer to one of its neighbors. With probability $q$ each node is *faulty*, independently of other nodes, in which case its advice points at an arbitrary neighbor, chosen uniformly at random. Otherwise, the node is *sound* and points at the correct neighbor. Crucially, the advice is *permanent*, in the sense that querying a node several times would yield the same answer. We evaluate efficiency by two measures: The *move complexity* denotes the expected number of edge traversals, and the *query complexity* denotes the expected number of queries.

Let $\Delta$ denote the maximal degree. Roughly speaking, the main message of this paper is that a phase transition occurs when the *noise parameterq* is roughly $1/\sqrt{\Delta}$. More precisely, we prove that above the threshold, every search algorithm has query complexity (and move complexity) which is both exponential in the depth $d$ of the treasure and polynomial in the number of nodes $n$. Conversely, below the threshold, there exists an algorithm with move complexity $O(d\sqrt{\Delta})$, and an algorithm with query complexity $O(\sqrt{\Delta}\log\Delta\log^2 n)$. Moreover, for the case of regular trees, we obtain an algorithm with query complexity $O(\sqrt{\Delta}\log n\log\log n)$. For $q$ that is below but close to the threshold, the bound for the move complexity is tight, and the bounds for the query complexity are not far from the lower bound of $\Omega(\sqrt{\Delta}\log_\Delta n)$.

In addition, we also consider a *semi-adversarial* variant, in which faulty nodes are still chosen at random, but an adversary chooses (beforehand) the advice of such nodes. For this variant, the threshold for efficient moving algorithms happens when the noise parameter is roughly $1/\Delta$. In fact, above this threshold a simple protocol that follows each advice with a fixed probability already achieves optimal move complexity.

### *7.1.3. Patterns on 3 vertices*

In [31] we deal with graph classes characterization and recognition. A popular way to characterize a graph class is to list a minimal set of forbidden induced subgraphs. Unfortunately this strategy usually does not lead to an efficient recognition algorithm. On the other hand, many graph classes can be efficiently recognized by techniques based on some interesting orderings of the nodes, such as the ones given by traversals.

We study specifically graph classes that have an ordering avoiding some ordered structures. More precisely, we consider what we call *patterns on three nodes*, and the recognition complexity of the associated classes. In this domain, there are two key previous works. Damashke started the study of the classes defined by forbidden patterns, a set that contains interval, chordal and bipartite graphs among others. On the algorithmic side, Hell, Mohar and Rafiey proved that any class defined by a set of forbidden patterns can be recognized in polynomial time. We improve on these two works, by characterizing systematically all the classes defined sets of forbidden patterns (on three nodes), and proving that among the 23 different classes (up to complementation) that we find, 21 can actually be recognized in linear time.

Beyond this result, we consider that this type of characterization is very useful, leads to a rich structure of classes, and generates a lot of open questions worth investigating.

### *7.1.4. The Dependent Doors Problem: An Investigation into Sequential Decisions without Feedback*

In [13], we introduce the *dependent doors problem* as an abstraction for situations in which one must perform a sequence of dependent decisions, without receiving feedback information on the effectiveness of previously made actions. Informally, the problem considers a set of $d$ doors that are initially closed, and the aim is to open all of them as fast as possible. To open a door, the algorithm knocks on it and it might open or not according to some probability distribution. This distribution may depend on which other doors are currently open, as well as on which other doors were open during each of the previous knocks on that door. The algorithm aims to minimize the expected time until all doors open. Crucially, it must act at any time without knowing whether or which other doors have already opened. In this work, we focus on scenarios where dependencies between doors are both positively correlated and acyclic.

The fundamental distribution of a door describes the probability it opens in the best of conditions (with respect to other doors being open or closed). We show that if in two configurations of $d$ doors corresponding doors share the same fundamental distribution, then these configurations have the same optimal running time up to a universal constant, no matter what are the dependencies between doors and what are the distributions. We also identify algorithms that are optimal up to a universal constant factor. For the case in which all doors share the same fundamental distribution we additionally provide a simpler algorithm, and a formula to calculate its running time. We furthermore analyse the price of lacking feedback for several configurations governed by standard fundamental distributions. In particular, we show that the price is logarithmic in $d$ for memoryless doors, but can potentially grow to be linear in $d$ for other distributions.

We then turn our attention to investigate precise bounds. Even for the case of two doors, identifying the optimal sequence is an intriguing combinatorial question. Here, we study the case of two cascading memoryless doors. That is, the first door opens on each knock independently with probability $p_1$. The second door can only open if the first door is open, in which case it will open on each knock independently with probability $p_2$. We solve this problem almost completely by identifying algorithms that are optimal up to an additive term of 1.

### *7.1.5. Finding maximum cliques in disk and unit ball graphs*

In an *intersection graph*, the vertices are geometric objects with an edge between any pair of intersecting objects. Intersection graphs have been studied for many different families of objects due to their practical applications and their rich structural properties. Among the most studied ones are *disk graphs*, which are intersection graphs of closed disks in the plane, and their special case, *unit disk graphs*, where all the radii are equal. Their applications range from sensor networks to map labeling, and many standard optimization problems have been studied on disk graphs. Most of the hard optimization and decision problems remain

NP-hard on disk graphs and even unit disk graphs. For instance, disk graphs contain planar graphs on which several of those problems are intractable.

The complexity of MAXIMUM CLIQUE on general disk graphs is a notorious open question in computational geometry. On the one hand, no polynomial-time algorithm is known, even when the geometric representation is given. On the other hand, the NP-hardness of the problem has not been established, even when only the graph is given as input.

Recently, Bonnet *et al.* showed that the disjoint union of two odd cycles is not the complement of a disk graph. From this result, they obtained a subexponential algorithm running in time $2^{\widetilde{O}(n^{2/3})}$ for MAXIMUM CLIQUE on disk graphs, based on a win-win approach. They also got a QPTAS by calling a PTAS for MAXIMUM INDEPENDENT SET on graphs with sublinear odd cycle packing number due to Bock *et al.*, or branching on a low-degree vertex.

In [17], our main contributions are twofold. The first is a randomized EPTAS (Efficient Polynomial-Time Approximation Scheme, that is, a PTAS in time $f(\varepsilon)n^{O(1)}$) for MAXIMUM INDEPENDENT SET on graphs of $\mathcal{X}(d, \beta, 1)$. The class $\mathcal{X}(d, \beta, 1)$ denotes the class of graphs whose neighborhood hypergraph has VC-dimension at most $d$, independence number at least $\beta n$, and no disjoint union of two odd cycles as an induced subgraph. Using the forbidden induced subgraph result of Bonnet *et al.*, it is then easy to reduce MAXIMUM CLIQUE on disk graphs to MAXIMUM INDEPENDENT SET on $\mathcal{X}(4, \beta, 1)$ for some constant $\beta$. We therefore obtain a randomized EPTAS (and a PTAS) for MAXIMUM CLIQUE on disk graphs, settling almost [0] completely the approximability of this problem.

The second contribution is to show the same forbidden induced subgraph for unit ball graphs as the one obtained for disk graphs : their complement cannot have a disjoint union of two odd cycles as an induced subgraph. The proofs are radically different and the classes are incomparable. So the fact that the same obstruction applies for disk graphs and unit ball graphs might be somewhat accidental. And again we therefore obtain a randomized EPTAS in time $2^{\widetilde{O}(1/\varepsilon^3)}n^{O(1)}$ for MAXIMUM CLIQUE on unit ball graphs, even without the geometric representation.

Before that result, the best approximation factor was 2.553, due to Afshani and Chan. In particular, even getting a 2-approximation algorithm (as for disk graphs) was open.

Finally we show that such an approximation scheme, even in subexponential time, is unlikely for ball graphs (that is, 3-dimensional disk graphs with arbitrary radii), and unit 4-dimensional disk graphs. Our lower bounds also imply NP-hardness. To the best of our knowledge, the NP-hardness of MAXIMUM CLIQUE on unit $d$-dimensional disk graphs was only known when $d$ is superconstant ($d = \Omega(\log n)$).

### 7.1.6. $\delta$-hyperbolicity

In [19], we show that the eccentricities (and thus the centrality indices) of all vertices of a $\delta$-hyperbolic graph $G = (V, E)$ can be computed in linear time with an additive one-sided error of at most $c\delta$, i.e., after a linear time preprocessing, for every vertex $v$ of $G$ one can compute in $O(1)$ time an estimate $\widehat{e}(v)$ of its eccentricity $ecc_G(v)$ such that $ecc_G(v) \leq \widehat{e}(v) \leq ecc_G(v) + c\delta$ for a small constant $c$. We prove that every $\delta$-hyperbolic graph $G$ has a shortest path tree, constructible in linear time, such that for every vertex $v$ of $G$, $ecc_G(v) \leq ecc_T(v) \leq ecc_G(v) + c\delta$. These results are based on an interesting monotonicity property of the eccentricity function of hyperbolic graphs: the closer a vertex is to the center of $G$, the smaller its eccentricity is. We also show that the distance matrix of $G$ with an additive one-sided error of at most $c'\delta$ can be computed in $O(|V|^2 \log^2 |V|)$ time, where $c' < c$ is a small constant. Recent empirical studies show that many real-world graphs (including Internet application networks, web networks, collaboration networks, social networks, biological networks, and others) have small hyperbolicity. So, we analyze the performance of our algorithms for approximating centrality and distance matrix on a number of real-world networks. Our experimental results show that the obtained estimates are even better than the theoretical bounds.

[0] The NP-hardness, ruling out a 1-approximation, is still to show.

### *7.1.7. Graph searches and geometric convexities in graphs*

In an attempt to understand graph searching on cocomparability graphs has been so successful, one quickly notices that the orderings produced by these traversals are precisely words of some antimatroids or convex geometries. The notion of antimatroids and convex geometries have appeared in the literature under various settings; in this work, we focus on the graph searching setting, where we discuss some known geometries on cocomparability graphs, and then present new structural properties on AT-free graphs in the hope of exploring whether the algorithms on cocomparability graphs can be lifted to this larger graph class. A first version of this work in collaboration with Feodor Dragan and Lalla Mouatadib was presented at ICGT Lyon, july 2018.

## 7.2. Distributed Computing

### *7.2.1. On the Limits of Noise in Distributed Computing*

Biological systems can share and collectively process information to yield emergent effects, despite inherent noise in communication. While man-made systems often employ intricate structural solutions to overcome noise, the structure of many biological systems is more amorphous. It is not well understood how communication noise may affect the computational repertoire of such groups. To approach this question we consider in [9], [15] the basic collective task of rumor spreading, in which information from few knowledgeable sources must reliably flow into the rest of the population. We study the effect of communication noise on the ability of groups that lack stable structures to efficiently solve this task. We present an impossibility result which strongly restricts reliable rumor spreading in such groups. Namely, we prove that, in the presence of even moderate levels of noise that affect all facets of the communication, no scheme can significantly outperform the trivial one in which agents have to wait until directly interacting with the sources—a process which requires linear time in the population size. Our results imply that in order to achieve efficient rumor spread a system must exhibit either some degree of structural stability or, alternatively, some facet of the communication which is immune to noise. We then corroborate this claim by providing new analyses of experimental data regarding recruitment in Cataglyphis niger desert ants. Finally, in light of our theoretical results, we discuss strategies to overcome noise in other biological systems.

### *7.2.2. Minimizing message size in stochastic communication patterns: fast self-stabilizing protocols with 3 bits*

In [8], we consider the basic PULL model of communication, in which in each round, each agent extracts information from few randomly chosen agents. We seek to identify the smallest amount of information revealed in each interaction (message size) that nevertheless allows for efficient and robust computations of fundamental information dissemination tasks. We focus on the *Majority Bit Dissemination* problem that considers a population of $n$ agents, with a designated subset of *source agents*. Each source agent holds an *input bit* and each agent holds an *output bit*. The goal is to let all agents converge their output bits on the most frequent input bit of the sources (the *majority bit*). Note that the particular case of a single source agent corresponds to the classical problem of *Broadcast* (also termed *Rumor Spreading*). We concentrate on the severe fault-tolerant context of *self-stabilization*, in which a correct configuration must be reached eventually, despite all agents starting the execution with arbitrary initial states. In particular, the specification of who is a source and what is its initial input bit may be set by an adversary.

We first design a general compiler which can essentially transform any self-stabilizing algorithm with a certain property (called "the *the bitwise-independence property*") that uses $\ell$-bits messages to one that uses only $\log \ell$-bits messages, while paying only a small penalty in the running time. By applying this compiler recursively we then obtain a self-stabilizing *Clock Synchronization* protocol, in which agents synchronize their clocks modulo some given integer $T$, within $\widetilde{O}(\log n \log T)$ rounds w.h.p., and using messages that contain 3 bits only. We then employ the new Clock Synchronization tool to obtain a self-stabilizing Majority Bit Dissemination protocol which converges in $\widetilde{O}(\log n)$ time, w.h.p., on every initial configuration, provided that the ratio of sources supporting the minority opinion is bounded away from half. Moreover, this protocol also uses only 3 bits per interaction.

### 7.2.3. Intense Competition can Drive Selfish Explorers to Optimize Coverage

In [30], we consider a game-theoretic setting in which selfish individuals compete over resources of varying quality. The motivating example is a group of animals that disperse over patches of food of different abundances. In such scenarios, individuals are biased towards selecting the higher quality patches, while, at the same time, aiming to avoid costly collisions or overlaps. Our goal is to investigate the impact of collision costs on the parallel coverage of resources by the whole group.

Consider $M$ *sites*, where a site $x$ has *value* $f(x)$. We think of $f(x)$ as the reward associated with site $x$, and assume that if a single individual visits $x$ exclusively, it receives this exact reward. Typically, we assume that if $\ell > 1$ individuals visit $x$ then each receives at most $f(x)/\ell$. In particular, when competition costs are high, each individual might receive an amount strictly less than $f(x)/\ell$, which could even be negative. Conversely, modeling cooperation at a site, we also consider cases where each one gets more than $f(x)/\ell$. There are $k$ identical *players* that compete over the rewards. They independently act in parallel, in a one-shot scenario, each specifying a single site to visit, without knowing which sites are explored by others. The group performance is evaluated by the expected *coverage*, defined as the sum of $f(x)$ over all sites that are explored by at least one player. Since we assume that players cannot coordinate before choosing their site we focus on symmetric strategies.

The main takeaway message of this paper is that the optimal symmetric coverage is expected to emerge when collision costs are relatively high, so that the following "Judgment of Solomon" type of rule holds: If a single player explores a site $x$ then it gains its full reward $f(x)$, but if several players explore it, then neither one receives any reward. Under this policy, it turns out that there exists a unique symmetric Nash Equilibrium strategy, which is, in fact, evolutionary stable. Moreover, this strategy yields the best possible coverage among all symmetric strategies. Viewing the coverage measure as the social welfare, this policy thus enjoys a *(Symmetric) Price of Anarchy* of precisely 1, whereas, in fact, any other congestion policy has a price strictly greater than 1.

Our model falls within the scope of mechanism design, and more precisely in the area of incentivizing exploration. It finds relevance in evolutionary ecology, and further connects to studies on Bayesian parallel search algorithms.

### 7.2.4. Universal Protocols for Information Dissemination Using Emergent Signals

In [23], we consider a population of $n$ agents which communicate with each other in a decentralized manner, through random pairwise interactions. One or more agents in the population may act as authoritative sources of information, and the objective of the remaining agents is to obtain information from or about these source agents. We study two basic tasks: broadcasting, in which the agents are to learn the bit-state of an authoritative source which is present in the population, and source detection, in which the agents are required to decide if at least one source agent is present in the population or not.

We focus on designing protocols which meet two natural conditions: (1) universality, i.e., independence of population size, and (2) rapid convergence to a correct global state after a reconfiguration, such as a change in the state of a source agent. Our main positive result is to show that both of these constraints can be met. For both the broadcasting problem and the source detection problem, we obtain solutions with a convergence time of $O(\log^2 n)$ rounds, w.h.p., from any starting configuration. The solution to broadcasting is exact, which means that all agents reach the state broadcast by the source, while the solution to source detection admits one-sided error on a $\varepsilon$-fraction of the population (which is unavoidable for this problem). Both protocols are easy to implement in practice and have a compact formulation.

Our protocols exploit the properties of self-organizing oscillatory dynamics. On the hardness side, our main structural insight is to prove that any protocol which meets the constraints of universality and of rapid convergence after reconfiguration must display a form of non-stationary behavior (of which oscillatory dynamics are an example). We also observe that the periodicity of the oscillatory behavior of the protocol, when present, must necessarily depend on the number $^{\#}X$ of source agents present in the population. For instance, our protocols inherently rely on the emergence of a signal passing through the population, whose

period is $\Theta(\log \frac{n}{\#X})$ rounds for most starting configurations. The design of clocks with tunable frequency may be of independent interest, notably in modeling biological networks.

### *7.2.5. Ergodic Effects in Token Circulation*

In [25], we consider a dynamical process in a network which distributes all particles (tokens) located at a node among its neighbors, in a round-robin manner.

We show that in the recurrent state of this dynamics (i.e., disregarding a polynomially long initialization phase of the system), the number of particles located on a given edge, averaged over an interval of time, is tightly concentrated around the average particle density in the system. Formally, for a system of $k$ particles in a graph of $m$ edges, during any interval of length $T$, this time-averaged value is $k/m \pm \widetilde{O}(1/T)$, whenever $gcd(m,k) = \widetilde{O}(1)$ (and so, e.g., whenever $m$ is a prime number). To achieve these bounds, we link the behavior of the studied dynamics to ergodic properties of traversals based on Eulerian circuits on a symmetric directed graph. These results are proved through sum set methods and are likely to be of independent interest.

As a corollary, we also obtain bounds on the *idleness* of the studied dynamics, i.e., on the longest possible time between two consecutive appearances of a token on an edge, taken over all edges. Designing trajectories for $k$ tokens in a way which minimizes idleness is fundamental to the study of the patrolling problem in networks. Our results immediately imply a bound of $\widetilde{O}(m/k)$ on the idleness of the studied process, showing that it is a distributed $\widetilde{O}(1)$-competitive solution to the patrolling task, for all of the covered cases. Our work also provides some further insights that may be interesting in load-balancing applications.

### *7.2.6. Improved Analysis of Deterministic Load-Balancing Schemes*

In [7], we consider the problem of deterministic load balancing of tokens in the discrete model. A set of $n$ processors is connected into a $d$-regular undirected network. In every time step, each processor exchanges some of its tokens with each of its neighbors in the network. The goal is to minimize the discrepancy between the number of tokens on the most-loaded and the least-loaded processor as quickly as possible.

Rabani et al. (1998) present a general technique for the analysis of a wide class of discrete load balancing algorithms. Their approach is to characterize the deviation between the actual loads of a discrete balancing algorithm with the distribution generated by a related Markov chain. The Markov chain can also be regarded as the underlying model of a continuous diffusion algorithm. Rabani et al. showed that after time $T = O(\log(Kn)/\mu)$, any algorithm of their class achieves a discrepancy of $O(d \log n/\mu)$, where $\mu$ is the spectral gap of the transition matrix of the graph, and $K$ is the initial load discrepancy in the system.

In this work we identify some natural additional conditions on deterministic balancing algorithms, resulting in a class of algorithms reaching a smaller discrepancy. This class contains well-known algorithms, eg., the Rotor-Router. Specifically, we introduce the notion of cumulatively fair load-balancing algorithms where in any interval of consecutive time steps, the total number of tokens sent out over an edge by a node is the same (up to constants) for all adjacent edges. We prove that algorithms which are cumulatively fair and where every node retains a sufficient part of its load in each step, achieve a discrepancy of $O(\min\{d\sqrt{\log n/\mu}, d\sqrt{n}\})$ in time $O(T)$. We also show that in general neither of these assumptions may be omitted without increasing discrepancy. We then show by a combinatorial potential reduction argument that any cumulatively fair scheme satisfying some additional assumptions achieves a discrepancy of $O(d)$ almost as quickly as the continuous diffusion process. This positive result applies to some of the simplest and most natural discrete load balancing schemes.

### *7.2.7. The assignment problem*

In the allocation problem, asynchronous processors must partition a set of items so that each processor leave knowing all items exclusively allocated to it. In [21], we introduce a new variant of the allocation problem called the assignment problem, in which processors might leave having only partial knowledge of their assigned items. The missing items in a processor's assignment must eventually be announced by other processors.

While allocation has consensus power 2, we show that the assignment problem is solvable read-write wait-free when $k$ processors compete for at least $2k-1$ items. Moreover, we propose a long-lived read-write wait-free assignment algorithm which is fair, allocating no more than 2 items per processor, and in which a slow processor may delay the assignment of at most $n$ items, where n is the number of processors.

The assignment problem and its read-write solution may be of practical interest for implementing resource allocators and work queues, which are pervasive concurrent programming patterns, as well as stream-processing systems.

### 7.2.8. A Characterization of $t$-Resilient Colorless Task Anonymous Solvability

One of the central questions in distributed computability is characterizing the tasks that are solvable in a given system model. In the anonymous case, where processes have no identifiers and communicate through multi-writer/multi-reader registers, there is a recent topological characterization (Yanagisawa 2017) of the colorless tasks that are solvable when any number of asynchronous processes may crash. In [22], we consider the case where at most $t$ asynchronous processes may crash, where $1 \leq t < n$. We prove that a colorless task is t-resilient solvable anonymously if and only if it is t-resilient solvable non-anonymously. We obtain our results through various reductions and simulations that explore how to extend techniques for non-anonymous computation to anonymous one.

### 7.2.9. Implementing Snapshot Objects on Top of Crash-Prone Asynchronous Message-Passing Systems

In asynchronous crash-prone read/write shared-memory systems there is the notion of a snapshot object, which simulates the behavior of an array of single-writer/multi-reader (SWMR) shared registers that can be read atomically. Processes in the system can access the object invoking (any number of times) two operations, denoted write() and snapshot(). A process invokes write() to update the value of its register in the array. When it invokes snapshot(), the process obtains the values of all registers, as if it read them simultaneously. It is known that a snapshot object can be implemented on top of SWMR registers, tolerating any number of process failures. Snapshot objects provide a level of abstraction higher than individual SWMR registers, and they simplify the design of applications. Building a snapshot object on an asynchronous crash-prone message-passing system has similar benefits. The object can be implemented by using the known simulations of a SWMR shared memory on top of an asynchronous message-passing system (if less than half the processes can crash), and then build a snapshot object on top of the simulated SWMR memory. [10] presents an algorithm that implements a snapshot object directly on top of the message-passing system, without building an intermediate layer of a SWMR shared memory. To the authors knowledge, the proposed algorithm is the first providing such a direct construction. The algorithm is more efficient than the indirect solution, yet relatively simple.

### 7.2.10. Distributed decision

We have carried out our study of distributed decision, either for its potential application to the design of fault-tolerant distributed algorithm, or for the purpose of designing a complexity/computanility theory for distributed network computing.

In the framework of *distributed network computing*, it is known that not all Turing-decidable predicates on labeled networks can be decided *locally* whenever the computing entities are Turing machines (TM), and this holds even if nodes are running *non-deterministic* Turing machines (NTM). In contrast, we show in [6] that every Turing-decidable predicate on labeled networks can be decided locally if nodes are running *alternating* Turing machines (ATM). More specifically, we show that, for every such predicate, there is a local algorithm for ATMs, with at most two alternations, that decides whether the actual labeled network satisfies that predicate. To this aim, we define a hierarchy of classes of decision tasks, where the lowest level contains tasks solvable with TMs, the first level those solvable with NTMs, and the level $k > 1$ contains those tasks solvable with ATMs with $k-1$ alternations. We characterize the entire hierarchy, and show that it collapses in the second level. In addition, we show separation results between the classes of network predicates that are locally decidable with TMs, NTMs, and ATMs, and we establish the existence of completeness results for

each of these classes, using novel notions of *local reduction*. We complete these results by a study of the local decision hierarchy when certificates are bounded to be of logarithmic size.

Distributed proofs are mechanisms enabling the nodes of a network to collectively and efficiently check the correctness of Boolean predicates on the structure of the network (e.g. having a specific diameter), or on data structures distributed over the nodes (e.g. a spanning tree). In [24], we consider well known mechanisms consisting of two components: a *prover* that assigns a *certificate* to each node, and a distributed algorithm called *verifier* that is in charge of verifying the distributed proof formed by the collection of all certificates. We show that many network predicates have distributed proofs offering a high level of redundancy, explicitly or implicitly. We use this remarkable property of distributed proofs to establish perfect tradeoffs between the *size of the certificate* stored at every node, and the *number of rounds* of the verification protocol.

The role of unique node identifiers in network computing is well understood as far as *symmetry breaking* is concerned. However, the unique identifiers also *leak information* about the computing environment—in particular, they provide some nodes with information related to the size of the network. It was recently proved that in the context of *local decision*, there are some decision problems that cannot be solved without unique identifiers, but unique identifiers leak a *sufficient* amount of information such that the problem becomes solvable (PODC 2013). In [11], we give a complete picture of what is the *minimal* amount of information that we need to leak from the environment to the nodes in order to solve local decision problems. Our key results are related to *scalar oracles* $f$ that, for any given $n$, provide a multiset $f(n)$ of $n$ labels; then the adversary assigns the labels to the $n$ nodes in the network. This is a direct generalisation of the usual assumption of unique node identifiers. We give a complete characterisation of the *weakest oracle* that leaks at least as much information as the unique identifiers. Our main result is the following dichotomy: we classify scalar oracles as *large* and *small*, depending on their asymptotic behaviour, and show that (1) any large oracle is at least as powerful as the unique identifiers in the context of local decision problems, while (2) for any small oracle there are local decision problems that still benefit from unique identifiers.

## 7.3. Models and Algorithms for Networks

### *7.3.1. Revisiting Radius, Diameter, and all Eccentricity Computation in Graphs through Certificates*

In [28], we introduce notions of certificates allowing to bound eccentricities in a graph. In particular , we revisit radius (minimum eccentricity) and diameter (maximum eccentricity) computation and explain the efficiency of practical radius and diameter algorithms by the existence of small certificates for radius and diameter plus few additional properties. We show how such computation is related to covering a graph with certain balls or complementary of balls. We introduce several new algorithmic techniques related to eccentricity computation and propose algorithms for radius, diameter and all eccentricities with theoretical guarantees with respect to certain graph parameters. This is complemented by experimental results on various real-world graphs showing that these parameters appear to be low in practice. We also obtain refined results in the case where the input graph has low doubling dimension, has low hyperbolicity, or is chordal.

### *7.3.2. Efficient Loop Detection in Forwarding Networks and Representing Atoms in a Field of Sets*

In [29], we consider the problem of detecting loops in a forwarding network which is known to be NP-complete when general rules such as wildcard expressions are used. Yet, network analyzer tools such as Netplumber (Kazemian et al., NSDI'13) or Veriflow (Khurshid et al., NSDI'13) efficiently solve this problem in networks with thousands of forwarding rules. In this paper, we complement such experimental validation of practical heuristics with the first provably efficient algorithm in the context of general rules. Our main tool is a canonical representation of the atoms (i.e. the minimal non-empty sets) of the field of sets generated by a collection of sets. This tool is particularly suited when the intersection of two sets can be efficiently computed and represented. In the case of forwarding networks, each forwarding rule is associated with the set of packet headers it matches. The atoms then correspond to classes of headers with same behavior in the network. We

propose an algorithm for atom computation and provide the first polynomial time algorithm for loop detection in terms of number of classes (which can be exponential in general). This contrasts with previous methods that can be exponential, even in simple cases with linear number of classes. Second, we introduce a notion of network dimension captured by the overlapping degree of forwarding rules. The values of this measure appear to be very low in practice and constant overlapping degree ensures polynomial number of header classes. Forwarding loop detection is thus polynomial in forwarding networks with constant overlapping degree.

### 7.3.3. Exact Distance Oracles Using Hopsets

In [33], we consider for fixed $h \geq 2$ the task of adding to a graph G a set of weighted shortcut edges on the same vertex set, such that the length of a shortest $h$-hop path between any pair of vertices in the augmented graph is exactly the same as the original distance between these vertices in G. A set of shortcut edges with this property is called an exact $h$-hopset and may be applied in processing distance queries on graph $G$. In particular, a 2-hopset directly corresponds to a distributed distance oracle known as a hub labeling. In this work, we explore centralized distance oracles based on 3-hopsets and display their advantages in several practical scenarios. In particular, for graphs of constant highway dimension, and more generally for graphs of constant skeleton dimension, we show that 3-hopsets require exponentially fewer shortcuts per node than any previously described distance oracle while incurring only a quadratic increase in the query decoding time, and actually offer a speedup when compared to simple oracles based on a direct application of 2-hopsets. Finally, we consider the problem of computing minimum-size h-hopset (for any $h \geq 2$) for a given graph $G$, showing a polylogarithmic-factor approximation for the case of unique shortest path graphs. When $h = 3$, for a given bound on the space used by the distance oracle, we provide a construction of hopsets achieving polylog approximation both for space and query time compared to the optimal 3-hopset oracle given the space bound.

### 7.3.4. Game Theory in Networks

Two notable contributions to game theory applied to networks are worth being mentioned.

In [14], we show that the Preferential Attachment rule naturally emerges in the context of evolutionary network formation, as the *unique* Nash equilibrium of a simple social network game. To demonstrate this result, we start from the fact that each node of a social network aims at maximizing its degree in the future, as this degree is representing its social capital in the "society" formed by the nodes and their connections. We show that, to maximize the node degree in the future, the unique Nash equilibrium consists in playing the Preferential Attachment rule when each node connects to the network. This result provides additional formal support to the commonly used Preferential Attachment model, initially designed to capture the "rich get richer" aphorism. In the process of establishing our result, we expose new connections between Preferential Attachment, random walks, and Young's Lattice.

In [20], we notice that distributed tasks such as constructing a maximal independent set (MIS) in a network, or properly coloring the nodes or the edges of a network with reasonably few colors, are known to admit efficient distributed randomized algorithms. Those algorithms essentially proceed according to some simple generic rules, by letting each node choosing a temptative value at random, and checking whether this choice is consistent with the choices of the nodes in its vicinity. If this is the case, then the node outputs the chosen value, else it repeats the same process. However, although such algorithms are, with high probability, running in a polylogarithmic number of rounds, they are *not robust* against actions performed by rational but selfish nodes. Indeed, such nodes may prefer specific individual outputs over others, e.g., because the formers suit better with some individual constraints. For instance, a node may prefer not being placed in a MIS as it is not willing to serve as a relay node. Similarly, a node may prefer not being assigned some radio frequencies (i.e., colors) as these frequencies would interfere with other devices running at that node. We show that the probability distribution governing the choices of the output values in the generic algorithm can be tuned such that no nodes will rationally deviate from this distribution. More formally, and more generally, we prove that the large class of so-called LCL tasks, including MIS and coloring, admit simple "Luby's style" algorithms where the probability distribution governing the individual choices of the output values forms a Nash equilibrium. In fact, we establish the existence of a stronger form of equilibria, called symmetric trembling-hand perfect equilibria for those games.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

### *8.1.1. Collaboration with Nokia Bell Labs*

Gang has a strong collaboration with Bell Labs (Nokia). We notably collaborate with Fabien Mathieu who is a former member of GANG and Élie de Panafieu. An ADR (joint research action) is dedicated to distributed learning.

This collaboration is developed inside the Alcatel-Lucent and Inria joint research lab.

# 9. Partnerships and Cooperations

## 9.1. Regional Initiatives

### *9.1.1. Laboratory of Information, Networking and Communication Sciences (LINCS)*

Gang is participating to the LINCS, a research centre co-founded by Inria, Institut Mines-Télécom, UPMC and Alcatel-Lucent Bell Labs, dedicated to research and innovation in the domains of future information and communication networks, systems and services. Gang contributes to work on online social networks, content centric networking and forwarding information verification.

## 9.2. National Initiatives

### *9.2.1. ANR DESCARTES*

**Participants:** Carole Delporte-Gallet, Hugues Fauconnier, Pierre Fraigniaud, Adrian Kosowski, Laurent Viennot.

Cyril Gavoille (U. Bordeaux) leads this project that grants 1 Post-Doc. H. Fauconnier is the local coordinator (This project began in October 2016).

Despite the practical interests of reusable frameworks for implementing specific distributed services, many of these frameworks still lack solid theoretical bases, and only provide partial solutions for a narrow range of services. We argue that this is mainly due to the lack of a generic framework that is able to unify the large body of fundamental knowledge on distributed computation that has been acquired over the last 40 years. The DESCARTES project aims at bridging this gap, by developing a systematic model of distributed computation that organizes the functionalities of a distributed computing system into reusable modular constructs assembled via well-defined mechanisms that maintain sound theoretical guarantees on the resulting system. DESCARTES arises from the strong belief that distributed computing is now mature enough to resolve the tension between the social needs for distributed computing systems, and the lack of a fundamentally sound and systematic way to realize these systems.

### *9.2.2. ANR MultiMod*

**Participants:** Adrian Kosowski, Laurent Viennot.

David Coudert (Sophia Antipolis) leads this project. L. Viennot coordinates locally. The project began in 2018.

The MultiMod project aims at enhancing the mobility of citizens in urban areas by providing them, through a unique interface enabling to express their preferences, the most convenient transportation means to reach their destinations. Indeed, the increasing involvement of actors and authorities in the deployment of more responsible and cost-effective logistics and the progress made in the field of digital technology have made possible to create synergies in the creation of innovative services for improving the mobility in cities. However, users are faced with a number of solutions that coexist at different scales, providing complementary information for the mobility of users, but that make very complex to find the most convenient itinerary at a given time for a specific user. In this context, MultiMod aims at improving the mobility of citizens in urban areas by proposing contextualized services, linking users, to facilitate multimodal transport by combining, with flexibility, all available modes (planned/dynamic carpooling, public transport (PT), car-sharing, bicycle, etc.).

We consider the use of carpooling in metropolitan areas, and so for short journeys. Such usage enables itineraries that are not possible with PT, allows for opening up areas with low PT coverage by bringing users near PT (last miles), and for faster travel-time when existing PT itineraries are too complex or with too low frequency (e.g., one bus per hour). In this context, the application must help the driver and the passenger as much as possible. In particular, the application must propose the meeting-point, indicate the driver the detour duration, and indicate the passenger how to reach this meeting-point using PT. Here, the time taken by drivers and passengers to agree becomes a critical issue and so the application must provide all needed information to quickly take a decision (i.e., in one click).

In addition, the era of Smart City gathers many emerging concepts, driven by innovative technological players, which enables the exploitation of real-time data (e.g., delay of a bus, traffic jam) made available by the various actors (e.g., communities in the framework of Open Data projects, users via their mobile terminals, traffic supervision authorities). In the MultiMod project, we will use these rich sources of data to propose itineraries that are feasible at query-time. Our findings will enable the design of a mobility companion able not only to guide the user along her journey, including when and how to change of transportation mean, but also to propose itinerary changes when the current one exceeds a threshold delay. The main originality of this project is thus to address the problem of computing itineraries in large-scale networks combining PT, carpooling and real-time data, and to satisfy the preferences of users. We envision that the outcome of this project will significantly improve the daily life of citizens.

The targeted metropolitan area for validating our solutions is Ile-de-France. Indeed, Instant-System is currently developing the new application “Vianavigo lab” which will replace the current “Vianavigo” application for the PT network of Ile-de-France. Our findings will therefore be tested at scale and eventually be integrated and deployed in production servers and mobile applications. The smaller networks of Bordeaux and Nice will be used to perform preliminary evaluations since Instant System already operates applications in these cities (Boogi Nice, Boogi Bordeaux). An important remark is that new features and algorithms can contractually be deployed in production every 4 months, thus enabling Instant System to measure and challenge the results of the MultiMod project in continue. This is a chance for the project to maximize its impact.

### *9.2.3. ANR FREDDA*

**Participants:** Carole Delporte-Gallet, Hugues Fauconnier, Pierre Fraigniaud.

Arnaud Sangnier (IRIF, Univ Paris Diderot) leads this project that grants 1 PhD. (This project began in October 2017).

Distributed algorithms are nowadays omnipresent in most systems and applications. It is of utmost importance to develop algorithmic solutions that are both robust and flexible, to be used in large scale applications. Currently, distributed algorithms are developed under precise assumptions on their execution context: synchronicity, bounds on the number of failures, etc. The robustness of distributed algorithms is a challenging problem that has not been much considered until now, and there is no systematic way to guarantee or verify the behavior of an algorithm beyond the context for which it has been designed. We propose to develop automated formal method techniques to verify the robustness of distributed algorithms and to support the development of robust applications. Our methods are of two kinds: statically through classical verification, and

dynamically, by synthesizing distributed monitors, that check either correctness or the validity of the context hypotheses at runtime.

### *9.2.4. ANR Distancia*

**Participants:** Pierre Charbit, Michel Habib, Laurent Viennot.

Victor Chepoi (Univ. Marseille) leads this project. P. Charbit coordinates locally. The project began in early-2018.

The theme of the project is Metric Graph Theory, and we are concerned both on theoretical foundations and applications. Such applications can be found in real world networks. For example, the hub labelling problem in road networks can be directly applied to car navigation applications. Understanding key structural properties of large-scale data networks is crucial for analyzing and optimizing their performance, as well as improving their reliability and security. In prior empirical and theoretical studies researchers have mainly focused on features such as small world phenomenon, power law degree distribution, navigability, and high clustering coefficients. Although those features are interesting and important, the impact of intrinsic geometric and topological features of large-scale data networks on performance, reliability and security is of much greater importance. Recently, there has been a surge of empirical works measuring and analyzing geometric characteristics of real-world networks, namely the Gromov hyperbolicity (called also the negative curvature) of the network. It has been shown that a number of data networks, including Internet application networks, web networks, collaboration networks, social networks, and others, have small hyperbolicity.

Metric graph theory was also indispensable in solving some open questions in concurrency and learning theory in computer science and geometric group theory in mathematics. Median graphs are exactly the 1–skeletons of CAT(0) cube complexes (which have been characterized by Gromov in a local-to-global combinatorial way). They play a vital role in geometric group theory (for example, in the recent solution of the famous Virtual Haken Conjecture). Median graphs are also the domains of event structures of Winskel, one of the basic abstract models of concurrency. This correspondence is very useful in dealing with questions on event structures.

Many classical algorithmic problems concern distances: shortest path, center and diameter, Voronoi diagrams, TSP, clustering, etc. Algorithmic and combinatorial problems related to distances also occur in data analysis. Low-distortion embeddings into l1-spaces (theorem of Bourgain and its algorithmical use by Linial et al.) were the founding tools in metric methods. Recently, several approximation algorithms for NP-hard problems were designed using metric methods. Other important algorithmic graph problems related to distances concern the construction of sparse subgraphs approximating inter-node distances and the converse, augmentation problems with distance constraints. Finally, in the distributed setting, an important problem is that of designing compact data structures allowing very fast computation of inter- node distances or routing along shortest or almost shortest paths. Besides computer science and mathematics, applications of structures involving distances can be found in archeology, computational biology, statistics, data analysis, etc. The problem of characterizing isometric subgraphs of hypercubes has its origin in communication theory and linguistics. . To take into account the recombination effect in genetic data, the mathematicians Bandelt and Dress developed in 1991 the theory of canonical decompositions of finite metric spaces. Together with geneticists, Bandelt successfully used it over the years to reconstruct phylogenies, in the evolutional analysis of mtDNA data in human genetics. One important step in their method is to build a reduced median network that spans the data but still contains all most parsimonious trees. As mentioned above, the median graphs occurring there constitute a central notion in metric graph theory.

With this project, we aim to participate at the elaboration of this new domain of Metric Graph Theory, which requires experts and knowledge in combinatorics (graphs, matroids), geometry, and algorithms. This expertise is distributed over the members of the consortium and a part of the success of our project it will be to share these knowledges among all the members of the consortium. This way we will create a strong group in France on graphs and metrics.

### 9.2.5. ANR HOSIGRA

**Participants:** Pierre Charbit, Michel Habib.

This project starting in early-2018, led by Reza Naserasr, explores the connection between minors and colorings, exploiting the notion of signed graphs. With the four colour theorem playing a central role in development of Graph Theory, the notions of minor and coloring have been branded as two of the most distinguished concepts in this field. The geometric notion of planarity has given birth to the theory of minors among others, and coloring have proven to have an algebraic nature through its extension to the theory of graph homomorphisms. Great many projects have been completed on both subjects, but what remains mostly a mystery is the correlation of the two subjects. The four color theorem itself, in slightly stronger form, claims that if a complete graph on five vertices cannot be formed by minor operation from a given graph, then the graph can be homomorphically mapped into the complete graph on four vertices (thus a 4-coloring). Commonly regarded as the most challenging conjecture on graph theory, the Hadwiger conjecture claims that five and four in this theorem can be replaced with $n$ and $n-1$ respectively for any value of $n$. The correlation of these two concepts has been difficult to study, mainly for the following reason: While the coloring or homomorphism problems roots back into intersections of odd-cycles, the minor operation is irrelevant of the parity of cycles. To overcome this barrier, the notion of signed graphs has been used implicitly since 1970s when coloring results on graphs with no odd-K4 is proved, following which a stronger form of the Hadwiger conjecture, known as Odd Hadwiger conjecture, was proposed by P. Seymour and B. Gerards, independently. Being a natural subclass of Matroids and a superclass of graphs, the notion of minor of signed graphs is well studied and many results from graph minor are either already extended to signed graphs or it is considered by experts of the subject. Observing the importance, and guided by some earlier works, in particular that of B. Guenin, we then started the study of algebraic concepts (coloring and homomormphisms) for signed graphs. Several results have been obtained in the past decade, and this project aims at exploring more of this topic.

## 9.3. European Initiatives

### 9.3.1. FP7 & H2020 Projects

Amos Korman has an ERC Consolidator Grant entitled "Distributed Biological Algorithms (DBA)", started in May 2015. This project proposes a new application for computational reasoning. More specifically, the purpose of this interdisciplinary project is to demonstrate the usefulness of an algorithmic perspective in studies of complex biological systems. We focus on the domain of collective behavior, and demonstrate the benefits of using techniques from the field of theoretical distributed computing in order to establish algorithmic insights regarding the behavior of biological ensembles. The project includes three related tasks, for which we have already obtained promising preliminary results. Each task contains a purely theoretical algorithmic component as well as one which integrates theoretical algorithmic studies with experiments. Most experiments are strategically designed by the PI based on computational insights, and are physically conducted by experimental biologists that have been carefully chosen by the PI. In turn, experimental outcomes will be theoretically analyzed via an algorithmic perspective. By this integration, we aim at deciphering how a biological individual (such as an ant) "thinks", without having direct access to the neurological process within its brain, and how such limited individuals assemble into ensembles that appear to be far greater than the sum of their parts. The ultimate vision behind this project is to enable the formation of a new scientific field, called algorithmic biology, that bases biological studies on theoretical algorithmic insights.

### 9.3.2. LIA Struco

Pierre Charbit is director of the LIA STRUCO, which is an Associated International Laboratory of CNRS between IÚUK, Prague, and IRIF, Paris. The director on the Czech side is Pr. Jaroslav Nešetřil. The primary theme of the laboratory is graph theory, more specifically: sparsity of graphs (nowhere dense classes of graphs, bounded expansion classes of graphs), extremal graph theory, graph coloring, Ramsey theory, universality and morphism duality, graph and matroid algorithms and model checking.

STRUCO focuses on high-level study of fundamental combinatorial objects, with a particular emphasis on comprehending and disseminating the state-of-the-art theories and techniques developed. The obtained insights shall be applied to obtain new results on existing problems as well as to identify directions and questions for future work.

One of the main goals of STRUCO is to provide a sustainable and reliable structure to help Czech and French researchers cooperate on long-term projects, disseminate the results to students of both countries and create links between these students more systematically. The chosen themes of the project indeed cover timely and difficult questions, for which a stable and significant cooperation structure is needed. By gathering an important number of excellent researchers and students, the LEA will create the required environment for making advances, which shall be achieved not only by short-term exchanges of researchers, but also by a strong involvement of Ph. D students in the learning of state-of-the-art techniques and in the international collaborations.

STRUCO is a natural place to federate and organize these many isolated collaborations between our two countries. Thus, the project would ensure long-term cooperations and allow young researchers (especially PhD students) to maintain the fruitful exchanges between the two countries in the future years, in a structured and federated way.

## 9.4. International Initiatives

### *9.4.1. Inria Associate Teams Not Involved in an Inria International Labs*

Carole Delporte-Gallet and Hugues Fauconnier are members of the Inria-MEXICO Equipe Associée LiDiCo (At the Limits of Distributed Computability, https://sites.google.com/site/lidicoequipeassociee/).

### *9.4.2. Inria International Partners*

*9.4.2.1. Informal International Partners*

Ofer Feinerman (Physics department of complex systems, Weizmann Institute of Science, Rehovot, Israel), is a team member in Amos Korman's ERC project DBA. This collaboration has been formally established by signing a contract between the CNRS and the Weizmann Institute of Science, as part of the ERC project.

Rachid Guerraoui (School of Computer and Communication Sciences, EPFL, Switzerland) maintains an active research collaboration with Gang team members (Carole Delporte, Hugues Fauconnier).

Sergio Rajsbaum (UNAM, Mexico) is a regular collaborator of the team, also involved formally in a joint French-Mexican research project (see next subsection).

Boaz Patt-Shamir (Tel Aviv University, Israel) is a regular collaborator of the team, also involved formally in a joint French-Israeli research project (see next subsection).

Lalla Moutadib, PhD student at Uinversity of Toronto, directed by Alan Borodin and Derek Corneil but also unformally by Michel Habib. 2 visits in 2018 in our group. She got her PhD in september 2018. See https://tspace.library.utoronto.ca/handle/1807/92081.

## 9.5. International Research Visitors

### *9.5.1. Visits of International Scientists*

- Sergio Rajsbaum ( UNAM Mexico) - April 1 to June 30.
- Giuliano Losa ( UCLA USA)- May 17 to May 30.

### *9.5.2. Visits to International Teams*

- Carole Delporte and Hugues Fauconnier have visited Sergio Rajsbaum at UNAM Mexico - September 2 to September 14.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

- Amos Korman: BDA 2018, General Chair of the organizing committee.
- Adrian Kosowski: WENDY Paris, chair of workshop.

*10.1.1.2. Member of the Organizing Committees*

Organisation of Dagstuhl Seminar 18211 *Formal Methods and Fault-Tolerant Distributed Computing: Forging an Alliance*, by Javier Esparza (TUM, Munick, Germany), Pierre Fraigniaud (IRIF and Inria GANG, Paris, France), Anca Muscholl (LaBRI, Bordeaux, France), and Sergio Rajsbaum (UNAM, Mexico, Mexique).

*10.1.1.3. Chair of Conference Program Committees*

- Amos Korman: BDA 2018, co-chair.

*10.1.1.4. Steering Committee Member*

- Amos Korman: ADGA 2018.
- Pierre Fraigniaud: *Highlights of Algorithms* (HALG) from January 2015.

*10.1.1.5. Member of the Conference Program Committees*

- Adrian Kosowski: MFCS 2018, SIROCCO 2018.
- Carole Delporte-Gallet: NETYS 2018.
- Pierre Fraigniaud: SPAA 2018, DISC 2018, ICALP 2018, WWW 2018, IPDPS 2018, LATIN 2018, HiPC 2018, ICDCN 2018.
- Michel Habib: WG 2018.

### *10.1.2. Journal*

*10.1.2.1. Member of the Editorial Boards*

- Pierre Fraigniaud is a member of the Editorial Board of Distributed Computing (DC).
- Pierre Fraigniaud is a member of the Editorial Board of Theory of Computing Systems (TOCS).
- Adrian Kosowski is a member of the Editorial Board of Mathematical Foundations of Computing (AIMS MFOC)

*10.1.2.2. Editors of a special issue*

- Carole Delporte is co-editors with Parosh Abdulla of the Special Issue on NETYS'2016 published in Computing ([27]).

### *10.1.3. Invited Talks*

- Hugues Fauconnier gives a seminar in College de France entitled "Failure detectors", December 2018.

### *10.1.4. Scientific Expertise*

- Adrian Kosowski was an expert panel member for grant panel PE6 of the National Science Center, Poland (Spring 2018).
- Pierre Fraigniaud was member of the *shadow committee* of the ERC Starting Grants selection panel in 2018.
- Pierre Fraigniaud was vice-president of the HCERES committee of Laboratoire d'Informatique de Polytechnique (LIX), November 2018.

### *10.1.5. Research Administration*

- Hugues Fauconnier is director of the UFR d'informatique of Université Paris Diderot.
- Carole Delporte-Gallet is deputy director of the UFR d'informatique of Université Paris Diderot.
- Laurent Viennot is leader of the "Algorithms and discrete structures" department of the Institute de Recherche en Informatique Fondamentale (IRIF).

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master: Carole Delporte and Hugues Fauconnier, Algorithmique distribuée avec mémoire partagée, 6h, M2, Université Paris Diderot

Master: Hugues Fauconnier, Cours programmation répartie, 33h, M2, Univ. Paris Diderot

Master: Carole Delporte, Cours et TP Protocoles des services internet, 44h, M2, Univ. Paris Diderot

Master: Carole Delporte, Cours Algorithmes répartis, 33h, M2, Univ. Paris Diderot

Master: Carole Delporte and Hugues Fauconnier, Théorie et pratique de la concurrence, 48h, M1, Université Paris Diderot

Licence: Carole Delporte and Hugues Fauconnier, Culture informatique, 16h, L2, Univ. Paris Diderot

Licence: Boufkhad Yacine, Algorithmique et Informatique, 132h, L1, IUT de l'Université Paris Diderot

Licence: Boufkhad Yacine, Programmation Orientée Objet, 60h, L2, IUT de l'Université Paris Diderot

Licence: Boufkhad Yacine, Traitement de données, 16h, L2, IUT de l'Université Paris Diderot

Master: Pierre Fraigniaud, Algorithmique parallèle et distribuée, 24h, Ecole Centrale Supélec Paris, M2

Master: Adrian Kosowski, Randomization in Computer Science: Games, Networks, Epidemic and Evolutionary Algorithms, 18h, M1, École Polytechnique

Licence: Adrian Kosowski, Design and Analysis of Algorithms, 32h, L3, École Polytechnique

Master: Pierre Fraigniaud and Adrian Kosowski, Algorithmique distribuée pour les réseaux, 24h, M2, Master Parisien de Recherche en Informatique (MPRI)

Master: Fabien de Montgolfier, Grand Réseaux d'Interaction, 44h, M2, Univ Paris Diderot

Licence: Fabien de Montgolfier, Protocoles Réseau (TP/TD), 24h, M1, Univ Paris Diderot

Licence: Fabien de Montgolfier, Programmation avancée (cours/TD/projet, bio-informatique), 52h, L3, Univ. Paris Diderot

Master: Fabien de Montgolfier, Algorithmique avancée (bio-informatique), 26h, M1, Univ Paris Diderot

Licence: Fabien de Montgolfier, Algorithmique (TD), 26h, L3, Ecole d'Ingénieurs Denis Diderot

Master : Laurent Viennot, Graph Mining, 6h, M2 MPRI, Univ. Paris Diderot

Licence: Pierre Charbit, Elements d'Algorithmique, 24h, L2, Université Paris Diderot, France

Licence: Pierre Charbit, Automates finis, 36h, L2, Université Paris Diderot, France

Licence: Pierre Charbit, Internet et Outils, 52h, L1, Université Paris Diderot, France

Master: Pierre Charbit, Programmation Objet, 60h, M2Pro PISE, Université Paris Diderot, France

Master: Pierre Charbit, Algorithmique de Graphes, 12h, M2 MPRI, Université Paris Diderot, France

### *10.2.2. Supervision*

PhD defended: Lucas Boczkowski (co-advised by Amos Korman and Iordanis Kerenidis). Title of thesis is: "Computing with Limited Resources in Uncertain Environments" [2]. Started September 2015, defended on November 30th, 2018.

PhD defended: Laurent Feuilloley (advised by Pierre Fraigniaud). Title of thesis is: "Synchronous Distributed Computing" [3]. Started September 2015, defended on September 19th, 2018.

PhD defended: Léo Planche (co-advised by Étienne Birmelé and Fabien de Montgolfier). Title of thesis is : "Graph Decomposition into Shortest Paths and Cycles of Small Eccentricity" [4]. Started October 2015, defended on November 23rd, 2018.

PhD defended: Vitaly Aksenov (co-advised by Petr Kuznetsov, Anatoly Shalyto and Carole Delporte). Title of thesis is : "Synchronization Costs in Parallel Programs and Concurrent Data Structures" [1]. Started October 2015, defended on September 26, 2018.

PhD in progress: Simon Collet (co-advised by Amos Korman and Pierre Fraigniaud). Title of thesis is: "Algorithmic Game Theory Applied to Biology". Started September 2015.

PhD in progress: Brieuc Guinard (advised by Amos Korman). Title of thesis is: "Algorithmic Aspects of Random Biological Processes". Started October 2016.

PhD in progress: Mengchuan Zou (co-advised by Adrian Kosowski and Michel Habib). Title of thesis is: "Local and Adaptive Algorithms for Optimization Problems in Large Networks". Started October 2016.

PhD in progress: Alkida Balliu and Dennis Olivetti (PhD students from L'Aquilla University and Gran Sasso Science Institute) are supervised by Pierre Fraigniaud.

PhD in progress: Lucas Hosseini (co-advised by Pierre Charbit, Patrice Ossona de Mendez and Jaroslav Nešetřil since Sept. 2014). Title : Limits of Structures.

Master internship (MPRI): Duc-Minh Phan (advised by Laurent Viennot). (March-August 2018) Title of report: "Public Transit Routing with Unrestricted Walking using Hub Labelling".

### *10.2.3. Juries*

Michel Habib was on the jury committee of the PhD thesis of Léo Planche: "Décomposition de graphes en plus courts chemins et en cycles de faible excentricité", Paris Descartes and Paris Diderot Universities, 23$^{th}$ novembre 2018.

Michel Habib was president of the jury committee of the PhD thesis of Julien Fradin : "Graphes complexes en biologie : problèmes, algorithmes et évaluations", Nantes University, 4$^{th}$ december 2018.

Michel Habib was on the jury committee of the PhD thesis of Mostafa Darwiche : "When operation research meets structural pattern recognition : on the solution of error-tolerant graph matching problems", Tours University, 5$^{th}$ december 2018.

Michel Habib was member of the jury for the HDR thesis of Jean-Sébastien Sereni : "Sur des aspects algébriques de la coloration de graphes : coloration fractionnaire et nombre de colorations", Université de Loraine, 13 février 2018.

Laurent Viennot was referee and on the jury committee of the HDR thesis of Frédéric Giroire on "Optimisation des infrastructures réseaux. Un peu de vert dans les réseaux et autres problèmes de placement et de gestion de ressources" at the University of Nice-Sophia Antipolis, October 2018.

Laurent Viennot was president of the jury committee of the PhD thesis of Matthieu Boutier "Routage sensible à la source" at Paris Diderot University, September 2018.

Laurent Viennot was on the jury committee of the PhD thesis of Alexandre Hollocou on "Novel Approaches to the Clustering of Large Graphs" at PSL University, December 2018.

Hugues Fauconnier was president of the jury committee of the PhD thesis of Vitaly Aksenov "Synchronization Costs in Parallel Programs and Concurrent Data Structures" at Paris Diderot University, September 2018.

Carole Delporte is on the jury committee of the PhD thesis of Vitaly Aksenov “Synchronization Costs in Parallel Programs and Concurrent Data Structures” at Paris Diderot University, September 2018.

Carole Delporte was president of the jury committee of the PhD thesis of Laurent Feuilloley “Local certification in distributed computing: error-sensitivity, uniformity, redundancy, and interactivity” at Paris Diderot University, September 2018.

Carole Delporte was referee and on the jury committee of the PhD thesis of Denis Jeanneau "Failure Detectors in Dynamic Distributed Systems" at Sorbonne université, December 2018.

Carole Delporte was president of the jury committee of the PhD thesis of Thibault Rieutord "Combinatorial Characterization of Asynchronous Distributed Computability" at Université Paris Saclay, Octobre 2018.

Pierre Fraigniaud was referee for the HDR thesis of Christine Tasson (IRIF, Paris Diderot) “Sémantiques vectorielles, probabilistes et distribuées”, 23 novembre 2018.

Pierre Fraigniaud was referee for the HDR thesis of Alessia Milani (LaBRI, Bordeaux) “Asynchronous Distributed Computing”, 12 novembre 2018.

Pierre Fraigniaud was member of the jury for the HDR thesis of Jean-Sébastien Sereni : “Sur des aspects algébriques de la coloration de graphes : coloration fractionnaire et nombre de colorations”, Université de Loraine, 13 février 2018.

Pierre Fraigniaud was member of the jury for the PhD thesis of Lucas Boczkowski “Distributed Computing Applied to Biology”, 30 novembre 2018.

## 10.3. Popularization

### *10.3.1. Internal or external Inria responsibilities*

- Laurent Viennot was “commissaire scientifique” for the permanent exposition on “Informatique et sciences du numérique” at Palais de la déécouverte in Paris (opened in March 2018).

### *10.3.2. Education*

- Carole Delporte was president of a jury of baccalaureat.

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] V. AKSENOV.*Synchronization Costs in Parallel Programs and Concurrent Data Structures*, ITMO University ; Paris Diderot University, September 2018, https://hal.inria.fr/tel-01887505

[2] L. BOCZKOWSKI.*Search and broadcast in stochastic environments, a biological perspective*, Université Paris 7, November 2018, https://tel.archives-ouvertes.fr/tel-01963290

[3] L. FEUILLOLEY.*Local certification in distributed computing: error-sensitivity, uniformity, redundancy, and interactivity*, Université paris diderot, September 2018, https://hal.archives-ouvertes.fr/tel-01962784

[4] L. PLANCHE.*Decomposing a graph with shortest paths of bounded eccentricity*, Paris Descartes, November 2018, https://tel.archives-ouvertes.fr/tel-01994139

## Articles in International Peer-Reviewed Journal

[5] P. ABOULKER, J. BANG-JENSEN, N. BOUSQUET, P. CHARBIT, F. HAVET, F. MAFFRAY, J. ZAMORA.*$\chi$-bounded families of oriented graphs*, in "Journal of Graph Theory", September 2018, vol. 89, n° 3, p. 304 - 326 [*DOI :* 10.1002/JGT.22252], https://hal.inria.fr/hal-01882395

[6] A. BALLIU, G. D 'ANGELO, P. FRAIGNIAUD, D. OLIVETTI.*What Can Be Verified Locally?*, in "Journal of Computer and System Sciences", 2018, https://hal.inria.fr/hal-01964764

[7] P. BERENBRINK, R. KLASING, A. KOSOWSKI, F. MALLMANN-TRENN, P. UZNANSKI.*Improved Analysis of Deterministic Load-Balancing Schemes*, in "ACM Transactions on Algorithms", December 2018, vol. 15, n° 1, Art.10 https://arxiv.org/abs/1404.4344 - minor corrections; updated literature overview [*DOI :* 10.1145/3282435], https://hal.inria.fr/hal-00979691

[8] L. BOCZKOWSKI, A. KORMAN, E. NATALE.*Minimizing message size in stochastic communication patterns: fast self-stabilizing protocols with 3 bits*, in "Distributed Computing", March 2018, https://hal.archives-ouvertes.fr/hal-01965945

[9] L. BOCZKOWSKI, E. NATALE, O. FEINERMAN, A. KORMAN.*Limits on reliable information flows through stochastic populations*, in "PLoS Computational Biology", June 2018, vol. 14, n° 6 [*DOI :* 10.1371/JOURNAL.PCBI.1006195], https://hal.inria.fr/hal-01953778

[10] C. DELPORTE-GALLET, H. FAUCONNIER, S. RAJSBAUM, M. RAYNAL.*Implementing Snapshot Objects on Top of Crash-Prone Asynchronous Message-Passing Systems*, in "IEEE Transactions on Parallel and Distributed Systems", September 2018, vol. 29, n° 9, p. 2033-2045 [*DOI :* 10.1109/TPDS.2018.2809551], https://hal.inria.fr/hal-01955906

[11] P. FRAIGNIAUD, J. HIRVONEN, J. SUOMELA.*Node labels in local decision*, in "Theoretical Computer Science", December 2018, vol. 751, p. 61-73, https://hal.inria.fr/hal-01964773

[12] M. HABIB, L. NOURINE.*Representation of lattices via set-colored posets*, in "Discrete Applied Mathematics", November 2018, vol. 249, p. 64-73 [*DOI :* 10.1016/J.DAM.2018.03.068], https://hal.inria.fr/hal-01955233

[13] A. KORMAN, Y. RODEH.*The Dependent Doors Problem: An Investigation into Sequential Decisions without Feedback*, in "ACM Transactions on Algorithms", October 2018, vol. 14, n° 4, p. 1-23, https://hal.inria.fr/hal-01953783

## International Conferences with Proceedings

[14] C. AVIN, A. COHEN, P. FRAIGNIAUD, Z. LOTKER, D. PELEG.*Preferential Attachment as a Unique Equilibrium*, in "World Wide Web Conference (WWW)", Lyon, France, 2018, p. 559-568 [*DOI :* 10.1145/3178876.3186122], https://hal.inria.fr/hal-01964759

[15] L. BOCZKOWSKI, O. FEINERMAN, A. KORMAN, E. NATALE.*Limits for Rumor Spreading in Stochastic Populations*, in "ITCS 2018 - 9th Innovations in Theoretical Computer Science Conference", Cambridge, United States, January 2018, p. 1-27, https://hal.archives-ouvertes.fr/hal-01958152

[16] L. BOCZKOWSKI, A. KORMAN, Y. RODEH.*Searching a Tree with Permanently Noisy Advice*, in "ESA 2018 - 26th Annual European Symposium on Algorithms", Helsinki, Finland, Leibniz International Proceedings

in Informatics (LIPIcs), August 2018, p. 1-32 [*DOI :* 10.4230/LIPIcs.ESA.2018.54], https://hal.archives-ouvertes.fr/hal-01958133

[17] M. Bonamy, E. Bonnet, N. Bousquet, P. Charbit, S. Thomassé.*EPTAS for Max Clique on Disks and Unit Balls*, in "FOCS: Foundations of Computer Science", Paris, France, October 2018, https://arxiv.org/abs/1803.01822 [*DOI :* 10.4230/LIPIcs], https://hal.inria.fr/hal-01962198

[18] E. Bonnet, N. Bousquet, P. Charbit, S. Thomassé, R. Watrigant.*Parameterized Complexity of Independent Set in H-free graphs*, in "IPEC 2018 - 13th International Symposium on Parameterized and Exact Computation", Helsinki, Finland, August 2018 [*DOI :* 10.4230/LIPIcs.CVIT.2016.23], https://hal.inria.fr/hal-01962369

[19] V. Chepoi, F. F. Dragan, M. Habib, Y. Vaxès, H. Alrasheed.*Fast Approximation of Centrality and Distances in Hyperbolic Graphs*, in "COCOA 2018 - 12th Annual International Conference on Combinatorial Optimization and Applications", Atlanta, United States, December 2018, p. 1-23, https://hal.inria.fr/hal-01955263

[20] S. Collet, P. Fraigniaud, P. Penna.*Equilibria of Games in Networks for Local Tasks*, in "22nd International Conference on Principles of Distributed Systems", Hong-Kong, China, 2018 [*DOI :* 10.4230/LIPIcs.OPODIS.2018.0], https://hal.inria.fr/hal-01964761

[21] C. Delporte-Gallet, H. Fauconnier, E. Gafni, G. Losa.*The Assignment Problem*, in "International Conference on Distributed Computing and Networking, ICDCN", Varanasi, India, 2018, https://hal.inria.fr/hal-01955902

[22] C. Delporte-Gallet, H. Fauconnier, S. Rajsbaum, N. Yanagisawa.*A Characterization of t-Resilient Colorless Task Anonymous Solvability*, in "Structural Information and Communication Complexity - 25th International Colloquium, SIROCCO", Ma'ale HaHamisha, Israel, 2018, https://hal.inria.fr/hal-01955837

[23] B. Dudek, A. Kosowski.*Universal Protocols for Information Dissemination Using Emergent Signals*, in "STOC 2018 Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing", Los Angeles, United States, ACM, 2018, p. 87-99, https://arxiv.org/abs/1705.09798 [*DOI :* 10.1145/3188745.3188818], https://hal.inria.fr/hal-01503359

[24] L. Feuilloley, P. Fraigniaud, J. Hirvonen, A. Paz, M. Perry.*Redundancy in Distributed Proofs*, in "32nd International Symposium on Distributed Computing", New Orleans, United States, 2018 [*DOI :* 10.4230/LIPIcs.DISC.2018.24], https://hal.inria.fr/hal-01964771

[25] A. Kosowski, P. Uznanski.*Ergodic Effects in Token Circulation*, in "SODA '18 Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms", New Orleans, United States, 2018, p. 2668-2682, https://hal.inria.fr/hal-01963249

### Conferences without Proceedings

[26] L. Boczkowski, B. Guinard, A. Korman, Z. Lotker, M. Renault.*Random Walks with Multiple Step Lengths*, in "LATIN 2018: Theoretical Informatics - 13th Latin American Symposium", Buenos Aires, Argentina, April 2018, p. 174-186, https://hal.archives-ouvertes.fr/hal-01955582

## Books or Proceedings Editing

[27] C. DELPORTE-GALLET, P. A. ABDULLA (editors). *Special Issue on NETYS 2016, Computing (journal)*, Springer, 2018, https://hal.inria.fr/hal-01955924

## Research Reports

[28] F. F. DRAGAN, M. HABIB, L. VIENNOT.*Revisiting Radius, Diameter, and all Eccentricity Computation in Graphs through Certificates*, Inria Paris ; Université paris diderot ; Kent State University, March 2018, https://arxiv.org/abs/1803.04660 , https://hal.inria.fr/hal-01729748

## Other Publications

[29] Y. BOUFKHAD, L. LINGUAGLOSSA, F. MATHIEU, D. PERINO, L. VIENNOT.*Efficient Loop Detection in Forwarding Networks and Representing Atoms in a Field of Sets*, September 2018, https://arxiv.org/abs/1809.01896 - working paper or preprint, https://hal.inria.fr/hal-01868778

[30] S. COLLET, A. KORMAN.*Intense Competition can Drive Selfish Explorers to Optimize Coverage*, December 2018, https://arxiv.org/abs/1805.01319 - working paper or preprint, https://hal.inria.fr/hal-01953800

[31] L. FEUILLOLEY, M. HABIB.*Graph classes and forbidden patterns on three vertices*, December 2018, https://arxiv.org/abs/1812.05913 - 40 pages, https://hal.inria.fr/hal-01958194

[32] P. FRAIGNIAUD, A. KORMAN, Y. RODEH.*Parallel Bayesian Search with no Coordination* , August 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01865469

[33] S. GUPTA, A. KOSOWSKI, L. VIENNOT.*Exact Distance Oracles Using Hopsets*, March 2018, https://arxiv.org/abs/1803.06977 - working paper or preprint, https://hal.inria.fr/hal-01737210

# Project-Team MAMBA

## Modelling and Analysis for Medical and Biological Applications

IN COLLABORATION WITH: Laboratoire Jacques-Louis Lions (LJLL)

IN PARTNERSHIP WITH:
**CNRS**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Modeling and Control for Life Sciences**

# Table of contents

# Project-Team MAMBA

*Creation of the Team: 2014 January 01, updated into Project-Team: 2015 April 01*

**Keywords:**

**Computer Science and Digital Science:**

A3. - Data and knowledge
A3.1. - Data
A3.1.1. - Modeling, representation
A3.4. - Machine learning and statistics
A3.4.6. - Neural networks
A3.4.7. - Kernel methods
A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.3. - Discrete Modeling (multi-agent, people centered)
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
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.2.6. - Optimization
A6.3. - Computation-data interaction
A6.3.1. - Inverse problems
A6.3.2. - Data assimilation
A6.4. - Automatic control
A6.4.1. - Deterministic control
A6.4.4. - Stability and Stabilization
A6.4.6. - Optimal control

**Other Research Topics and Application Domains:**

B1. - Life sciences
B1.1. - Biology
B1.1.2. - Molecular and cellular biology
B1.1.5. - Immunology
B1.1.6. - Evolutionnary biology
B1.1.7. - Bioinformatics
B1.1.8. - Mathematical biology
B1.2. - Neuroscience and cognitive science
B2. - Health
B2.2. - Physiology and diseases

B2.2.3. - Cancer
B2.2.4. - Infectious diseases, Virology
B2.2.6. - Neurodegenerative diseases
B2.3. - Epidemiology
B2.4. - Therapies
B2.4.1. - Pharmaco kinetics and dynamics
B2.4.2. - Drug resistance
B2.6.3. - Biological Imaging
B9.6.4. - Management science

# 1. Team, Visitors, External Collaborators

**Research Scientists**
Marie Doumic [Team leader, Inria, Senior Researcher, HDR]
Pierre-Alexandre Bliman [Inria, Senior Researcher, HDR]
Jean Clairambault [Inria, Senior Researcher, HDR]
Dirk Drasdo [Inria, Senior Researcher, HDR]
Luis Lopes Neves de Almeida [CNRS, Senior Researcher, HDR]
Diane Peurichard [Inria, Researcher]
Nastassia Pouradier Duteil [Inria, Researcher, from Sep 2018]
Philippe Robert [Inria, Senior Researcher, HDR]

**Faculty Members**
Stephane Mischler [Univ de Dauphine, Professor, until Aug 2018]
Ayman Moussa [Univ Pierre et Marie Curie, Associate Professor, from Sep 2018]
Benoît Perthame [Univ Pierre et Marie Curie, Professor, HDR]

**External Collaborators**
Noémie Boissier [Inria/IfADo]
Jieling Zhao [IfADO, Dortmund, Germany]

**Technical Staff**
Florian Joly [Inria, from Aug 2018]
Thais Roque [Inria, Oct 2018]
Paul Van Liedekerke [Inria]
Yi Yin [Inria, until Feb 2018]

**PhD Students**
Federica Bubba [Univ Pierre et Marie Curie]
Julia Delacour [Ecole Normale Supérieure Lyon]
Cecile Della Valle [Inria, from Sep 2018]
Hugo Martin [Univ Pierre et Marie Curie]
Mathieu Mézache [Inria]
Camille Pouchol [Univ Pierre et Marie Curie, until Aug 2018]
Alexandre Poulain [ERC, from Sep 2018]
Martin Strugarek [Ecole Nationale des Ponts et Chaussées, until Aug 2018]
Wen Sun [Inria, until Sep 2018]
Guilherme Thompson [Inria, granted by Brazilian Grant, until Aug 2018]
Gaëtan Vignoud [Ecole Normale Supérieure, from Sep 2018]

**Post-Doctoral Fellows**
Cécile Carrère [Univ Pierre et Marie Curie]
Davit Martirosyan [Inria, until Aug 2018]

Thais Roque [Inria, from May 2018 to Oct 2018]
Xinran Ruan [Inria]

**Administrative Assistant**
Nicole Moussa [Inria, from Apr 2018]

# 2. Overall Objectives

## 2.1. Context and overall objectives of the project-team

The MAMBA (Modelling and Analysis in Medical and Biological Applications) team is the continuation of the BANG (Biophysics, Numerical Analysis and Geophysics) team, which itself was a continuation of the former project-team M3N. Historically, the BANG team, headed by Benoît Perthame during 11 years (2003-2013), has developed models, simulations and numerical algorithms for problems involving dynamics of Partial Differential Equations (PDEs).

The dynamics of complex physical or biophysical phenomena involves many agents, e.g. proteins or cells - which can be seen as active agents. Mathematically, they can be represented either explicitly as individuals with their dynamics modelled e.g. through branching trees and piecewise deterministic Markov processes (PDMP), or stochastic differential equations, or under certain conditions be grouped or locally averaged, in which case their dynamics is mimicked by Ordinary or Partial Differential Equations (ODEs/PDEs).

Biology and medicine presently face the difficulty to make sense of the data newly available by means of recent signal acquisition methods and to take appropriate actions through possible treatment pathways. Modeling through agent-based or continuous models is a unique way to explain (model) the observations and then compute, control and predict the consequences of the mechanisms under study. These are the overall goals of Mamba.

# 3. Research Program

## 3.1. Introduction

Data and image analysis, statistical, ODEs, PDEs, and agent-based approaches are used either individually or in combination, with a strong focus on PDE analysis and agent-based approaches. Mamba was created in January 2014 as a continuation of the BANG project-team, that had been headed by Benoît Perthame from 2003-2013, and in the last years increasingly broadened its subjects as its members developed their own research agendas. It aims at developing models, simulations, numerical and control algorithms to solve questions from life sciences involving dynamics of phenomena encountered in biological systems such as protein intracellular spatio-temporal dynamics, cell motion, early embryonic development, multicellular growth, wound healing and liver regeneration, cancer evolution, healthy and tumor growth control by pharmaceuticals, protein polymerization occurring in neurodegenerative disorders, control of dengue epidemics, etc.

Another guideline of our project is to remain close to the most recent questions of experimental biology or medicine, to design models and problems under study as well as the related experiments to be carried out by our collaborators in biology or medicine. In this context, our ongoing collaborations with biologists and physicians: the collaboration with St Antoine Hospital in Paris within the Institut Universitaire de Cancérologie of Sorbonne Université (IUC, Luis Almeida, Jean Clairambault, Dirk Drasdo, Alexander Lorz, Benoît Perthame); Institut Jacques Monod (Luis Almeida); the INRA team headed by Human Rezaei and Wei-Feng Xue's team in the university of Canterbury through the ERC Starting Grant SKIPPER$^{AD}$ (Marie Doumic); our collaborators within the HTE program (François Delhommeau at St Antoine, Thierry Jaffredo, and Delphine Salort at IBPS, Sorbonne Université, Paris; François Vallette at INSERM Nantes); Frédéric Thomas at CREEC, Montpellier; Hôpital Paul Brousse through ANR-IFlow and ANR-iLite; and the close experimental collaborations that emerged through the former associate team QUANTISS (Dirk Drasdo), particularly at the Leibniz Institute for Working Environment and Human Factors in Dortmund, Germany; or more recently with Yves Dumont at CIRAD, Montpellier, are key points in our project.

Our main objective is the creation, investigation and transfer of new models, methods (for analysis but also for control) and algorithms. In selected cases software development as that of CellSys and TiQuant by D. Drasdo and S. Hoehme is performed. More frequently, the team develops "proof of concept" numerical codes in order to test the adequacy of our models to experimental biology.

Taking advantage of the last 4-year evaluation of MAMBA (September 2017), we have reorganized the presentation of our research program in three main methodological axes. Two main application axes are presented in the next Section. Evolving along their own logic in close interaction with the methodological axes, they are considered as application-driven research axes in themselves. The methodological research axes are the following.

*Axis 1* is devoted to work in physiologically-based design, analysis and control of population dynamics. It encompasses populations of bacteria, of cancer cells, of neurons, of aggregating proteins, etc. whose dynamics are represented by partial differential equations (PDEs), structured in evolving physiological traits, such as cell age, cell size, time elapsed since last firing (neurons).

*Axis 2* is devoted to reaction equations and motion equations of agents in living systems. It aims at describing biological phenomena such as tumor growth, chemotaxis and wound healing.

*Axis 3* tackles the question of model and parameter identification, combining stochastic and deterministic approaches and inverse problem methods in nonlocal and multi-scale models.

## 3.2. Methodological axis 1: analysis and control for population dynamics

**Personnel**

Pierre-Alexandre Bliman, Jean Clairambault, Marie Doumic, Benoît Perthame, Nastassia Pouradier Duteil, Philippe Robert

**Project-team positioning**

Population dynamics is a field with varied and wide applications, many of them being in the core of MAMBA interests - cancer, bacterial growth, protein aggregation. Their theoretical study also brings a qualitative understanding on the interplay between individual growth, propagation and reproduction in such populations. In the previous periods of evaluation, many results where obtained in the BANG team on the asymptotic and qualitative behavior of such structured population equations, see e.g. [126], [73], [94], [84]. Other Inria teams interested by this domain are Mycenae, Numed and Dracula, with which we are in close contacts. Among the leaders of the domain abroad, we can cite among others our colleagues Tom Banks (USA), Graeme Wake (New Zealand), Glenn Webb (USA), Jacek Banasiak (South Africa), Odo Diekmann (Netherlands), with whom we are also in regular contact. Most remarkably and recently, connections have also been made with probabilists working on Piecewise Deterministic Markov Processes (F. Malrieu at the university of Rennes, Jean Bertoin at the ETH in Zurich, Vincent Bansaye at Ecole Polytechnique, Julien Berestycki at Cambridge, Amaury Lambert at College de France, M. Hoffmann at Paris Dauphine), leading to a better understanding of the links between both types of results – see also the Methodological axis 3.

**Scientific achievements**

We divide this research axis, which relies on the study of structured population equations, according to four different applications, bringing their own mathematical questions, e.g., stability, control, or blow-up.

**Time asymptotics for nucleation, growth and division equations**

Following the many results obtained in the BANG team on the asymptotic and qualitative behavior of structured population equation, we put our effort on the investigation of limit cases, where the trend to a steady state or to a steady exponential growth described by the first eigenvector fails to happen. In [78], the case of equal mitosis (division into two equally-sized offspring) with linear growth rate was studied, and strangely enough, it appeared that the general relative entropy method could also be adapted to such a non-dissipative case. Many discussions and common workshops with probabilists, especially through the ANR project PIECE coordinated by F. Malrieu, have led both communities to work closer.

In [92], the case of constant fragmentation rate and linear growth rate has been investigated in a deterministic approach, whereas similar questions were simultaneously raised but in a stochastic process approach in [75].

We also enriched the models by taking into account a nucleation term, modeling the spontaneous formation of large polymers out of monomers [137]. We investigated the interplay between four processes: nucleation, polymerization, depolymerization and fragmentation.

The ERC Starting Grant SKIPPER$^{AD}$ (Doumic) supported and was the guideline for the study of nucleation, growth and fragmentation equations.

**Cell population dynamics and its control**
One of the important incentives for such model design, source of many theoretical works, is the challenging question of drug-induced drug resistance in cancer cell populations, described in more detail below in the Applicative axis 1, Cancer. The adaptive dynamics setting used consists of phenotype-structured integro-differential [or reaction-diffusion, when phenotype instability is added under the form of a Laplacian] equations describing the dynamic behavior of different cell populations interacting in a Lotka-Volterra-like manner that represents common growth limitation due to scarcity of expansion space and nutrients. The phenotype structure allows us to analyse the evolution in phenotypic traits of the populations under study and its asymptotics for two populations [119], [116], [115], [117]. Space may be added as a complementary structure variable provided that something is known of the (Cartesian) geometry of the population [118], which is seldom the case.

**Modelling, observation and identification of the spread of infectious diseases**
Epidemiological models are made to understand and predict the dynamics of the spread of infectious diseases. We initiated studies with the aim to understand how to use epidemiological data (typically given through incidence rate) in order to estimate the state of the population as well as constants, characteristic of the epidemics such as the transmission rate. The methods rely on observation and identification techniques borrowed from control theory.

**Modelling Mendelian and non-Mendelian inheritances in density-dependent population dynamics**
Classical strategies for controlling mosquitoes responsible of vector-borne disease are based on mechanical methods, such as elimination of oviposition sites; and chemical methods, such as insecticide spraying. Long term usage of the latter generates resistance [81], [103], transmitted to progeny according to Mendelian inheritance (in which each parent contributes randomly one of two possible alleles for a trait). New control strategies involve biological methods such as genetic control, which may either reduces mosquito population in a specific area or decreases the mosquito vector competence [61], [112], [144]. Among the latter, infection of wild populations by the bacterium *Wolbachia* appears promising (see also Applicative axis 2 below). Being maternally-transmitted, the latter obeys non-Mendelian inheritance law. Motivated by the effects of the (possibly unwanted) interaction of these two types of treatment, we initiated the study of modelling of Mendelian and non-Mendelian inheritances in density-dependent population dynamics.

**Control of collective dynamics**
The term *self-organization* is used to describe the emergence of complex organizational patterns from simple interaction rules in collective dynamics systems. Such systems are valuable tools to model various biological systems or opinion dynamics, whether it be the collective movement of animal groups, the organization of cells in an organism or the evolution of opinions in a large crowd. A special case of self-organization is given by *consensus*, i.e. the situation in which all agents' state variables converge. Another phenomenon is that of *clustering*, when the group is split into clusters that each converge to a different state. We have designed optimal control strategies to drive collective dynamics to consensus. In the case where consensus and clustering are situations to be avoided (for example in crowd dynamics), we designed control strategies to keep the system away from clustering.

**Models of neural network**

Mean field limits have been proposed by biophysicists in order to describe neural networks based on physiological models. The various resulting equations are called integrate-and-fire, time elapsed models, voltage-conductance models. Their specific nonlinearities and the blow-up phenomena make their originality which has led to develop specific mathematical analysis [129], followed by [124], [111], [130], [83]. This field also yields a beautiful illustration for the capacity of the team to combine and compare stochastic and PDE modelling (see Methodological axis 3), in [87].

**Models of interacting particle systems**

The organisation of biological tissues during development is accompanied by the formation of sharp borders between distinct cell populations. The maintenance of this cell segregation is key in adult tissue homeostatis, and its disruption can lead tumor cells to spread and form metastasis. This segregation is challenged during tissue growth and morphogenesis due to the high mobility of many cells that can lead to intermingling. Therefore, understanding the mechanisms involved in the generation and maintain of cell segregation is of tremendous importance in tissue morphogenesis, homeostasis, and in the development of various invasive diseases such as tumors. In this research axis, we aim to provide a mathematical framework which enables to quantitatively link the segregation and border sharpening ability of the tissue to these cell-cell interaction phenomena of interest [72]. As agent-based models do not enable precise mathematical analysis of their solutions due to the lack of theoretical results, we turn towards continuous -macroscopic- models and aim to provide a rigorous link between the different models [71].

**Collaborations**

- Nucleation, growth and fragmentation equations: **Juan Calvo**, university of Granada, came for two one-month visits, **Miguel Escobedo**, University of Bilbao (see also Methodological axis 3), **Pierre Gabriel**, University of Versailles-Saint Quentin, former B. Perthame and M. Doumic's Ph.D student, who now co-supervises Hugo Martin's Ph.D thesis. **Piotr Gwiązda**, Polish Academy of Sciences, Poland, **Emil Wiedemann**, University of Bonn, Germany, **Klemens Fellner**, university of Graz, Austria.
- Cell population dynamics and its control: **Tommaso Lorenzi**, former Mamba postdoc, now at the University of St. Andrews, Scotland, maintains a vivid collaboration with the Mamba team. He is in particular an external member of the HTE program MoGlImaging (see also Applicative axis 1). **Emmanuel Trélat**, Sorbonne Université professor, member of LJLL and of the CAGE Inria team, is the closest Mamba collaborator for optimal control. **Benedetto Piccoli**, Professor at Rutgers University (Camden, New Jersey), is collaborating on the analysis and control of collective dynamics.
- Mendelian inheritance and resistance in density-dependent population dynamics: **Pastor Pérez-Estigarribia, Christian Schaerer**, Universidad Nacional de Asunción, Paraguay.
- Neural networks: **Delphine Salort,** Professor Sorbonne Université, Laboratory for computations and quantification in biology, and **Patricia Reynaud**, University of Nice, **Maria Cáceres**, University of Granada.
- Models of interacting particle systems: **Pierre Degond**, Imperial College London, **MAPMO, Orléans**, **Ewelina Zatorska**, University College London, **Anais Khuong**, Francis Crick Institute

## 3.3. Methodological axis 2: reaction and motion equations for living systems

**Personnel**

Luis Almeida, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil.

**Project-team positioning**

The Mamba team had initiated and is a leader on the works developed in this research axis. It is a part of a consortium of several mathematicians in France through the ANR Blanc project *Kibord*, which involves in particular members from others Inria team (DRACULA, REO). Finally, we mention that from Sept. 2017 on, Mamba benefited from the ERC Advanced Grant ADORA (Asymptotic approach to spatial and dynamical organizations) of Benoît Perthame.

**Scientific achievements**

We divide this research axis, which relies on the study of partial differential equations for space and time organisation of biological populations, according to various applications using the same type of mathematical formalisms and methodologies: asymptotic analysis, weak solutions, numerical algorithms.

**Aggregation equation**

In the mathematical study of collective behavior, an important class of models is given by the aggregation equation. In the presence of a non-smooth interaction potential, solutions of such systems may blow up in finite time. To overcome this difficulty, we have defined weak measure-valued solutions in the sense of duality and its equivalence with gradient flows and entropy solutions in one dimension [109]. The extension to higher dimensions has been studied in [86]. An interesting consequence of this approach is the possibility to use the traditional finite volume approach to design numerical schemes able to capture the good behavior of such weak measure-valued solutions [102], [108].

**Identification of the mechanisms of single cell motion.**

In this research axis, we aim to study the mechanisms of single cell adhesion-based and adhesion free motion. This work is done in the frame of the recently created associated team MaMoCeMa (see Section 9) with the WPI, Vienna. In a first direction [140] with N. Sfakianakis (Heidelberg University), we extended the live-cell motility Filament Based Lamellipodium Model to incorporate the forces exerted on the lamellipodium of the cells due to cell-cell collision and cadherin induced cell-cell adhesion. We took into account the nature of these forces via physical and biological constraints and modelling assumptions. We investigated the effect these new components had in the migration and morphology of the cells through particular experiments. We exhibit moreover the similarities between our simulated cells and HeLa cancer cells.

In a second work done in collaboration with the group of biologist at IST (led by **Michael Sixt** Austria), we developed and analyzed a two-dimensional mathematical model for cells migrating without adhesion capabilities [110]. Cells are represented by their cortex, which is modelled as an elastic curve, subject to an internal pressure force. Net polymerization or depolymerization in the cortex is modelled via local addition or removal of material, driving a cortical flow. The model takes the form of a fully nonlinear degenerate parabolic system. An existence analysis is carried out by adapting ideas from the theory of gradient flows. Numerical simulations show that these simple rules can account for the behavior observed in experiments, suggesting a possible mechanical mechanism for adhesion-independent motility.

**Free boundary problems for tumor growth.**

Fluid dynamic equations are now commonly used to describe tumor growth with two main classes of models: those which describe tumor growth through the dynamics of the density of tumoral cells subjected to a mechanical stress; those describing the tumor through the dynamics of its geometrical domain thanks to a Hele-Shaw-type free boundary model. The first link between these two classes of models has been rigorously obtained thanks to an incompressible limit in [128] for a simple model. This result has motivated the use of another strategy based on viscosity solutions, leading to similar results, in [113].

Since more realistic systems are used in the analysis of medical images, we have extended these studies to include active motion of cells in [127], viscosity in [132] and proved regularity results in [120]. The limiting Hele-Shaw free boundary model has been used to describe mathematically the invasion capacity of a tumour by looking for travelling wave solutions, in [131], see also Methodological axis 3. It is a fundamental but difficult issue to explain rigorously the emergence of instabilities in the direction transversal to the wave propagation. For a simplified model, a complete explanation is obtained in [114].

**Two-way coupling of diffusion and growth.**

We are currently developing a mathematical framework for diffusion equations on time-evolving manifolds, where the evolution of the manifold is a function of the distribution of the diffusing quantity. The need for such a framework takes it roots in developmental biology. Indeed, the growth of an organism is triggered by signaling molecules called morphogens that diffuse in the organism during its development. Meanwhile, the diffusion of the morphogens is itself affected by the changes in shape and size of the organism. In other words, there is a complete coupling between the diffusion of the morphogens and the evolution of the shapes. In addition to the elaboration of this theoretical framework, we also collaborate with a team of developmental biologists from Rutgers University (Camden, New Jersey) to develop a model for the diffusion of Gurken during the oogenesis of *Drosophila*.

**Collaborations**

- Shanghai Jiao Tong University, joint publications with Min Tang on bacterial models for chemotaxis and free boundary problems for tumor growth.
- Imperial College London, joint works with José Antonio Carrillo on aggregation equation.
- University of Maryland at College Park, UCLA, Univ. of Chicago, Univ. Autónoma de Madrid, Univ. of St. Andrews (Scotland), joint works on mathematics of tumor growth models.
- Joint work with Francesco Rossi (Università di Padova, Italy) and Benedetto Piccoli (Rutgers University, Camden, New Jersey, USA) on Developmental PDEs.
- Cooperation with Shugo Yasuda (University of Hyogo, Kobe, Japan) and Vincent Calvez (EPI Dracula) on the subject of bacterial motion.
- Cooperation with Nathalie Ferrand (INSERM), Michèle Sabbah (INSERM) and Guillaume Vidal (Centre de Recherche Paul Pascal, Bordeaux) on cell aggregation by chemotaxis.
- Nicolas Vauchelet, Université Paris 13

## 3.4. Methodological axis 3: Model and parameter identification combining stochastic and deterministic approaches in nonlocal and multi-scale models

**Personnel**

Marie Doumic, Dirk Drasdo.

**Project-team positioning**

Mamba developed and addressed model and parameter identification methods and strategies in a number of mathematical and computational model applications including growth and fragmentation processes emerging in bacterial growth and protein misfolding, in liver regeneration [97], TRAIL treatment of HeLa cells [74], growth of multicellular spheroids [107], blood detoxification after drug-induced liver damage [139], [101].

This naturally led to increasingly combine methods from various fields: image analysis, statistics, probability, numerical analysis, PDEs, ODEs, agent-based modeling methods, involving inverse methods as well as direct model and model parameter identification in biological and biomedical applications. Model types comprise agent-based simulations for which Mamba is among the leading international groups, and Pharmocokinetic (PK) simulations that have recently combined in integrated models (PhD theses Géraldine Cellière, Noémie Boissier). The challenges related with the methodological variability has led to very fruitful collaborations with internationally renowned specialists of these fields, e.g. for bacterial growth and protein misfolding with Marc Hoffmann (Paris Dauphine) and Patricia Reynaud-Bouret (University of Nice) in statistics, with Tom Banks (Raleigh, USA) and Philippe Moireau (Inria M3DISIM) in inverse problems and data assimilation, and with numerous experimentalists.

**Scientific achievements**

Direct parameter identification is a great challenge particularly in living systems in which part of parameters at a certain level are under control of processes at smaller scales.

**Estimation methods for growing and dividing populations**
In this domain, all originated in two papers in collaboration with J.P. Zubelli in 2007 [133], [96], whose central idea was to used the asymptotic steady distribution of the individuals to estimate the division rate. A series of papers improved and extended these first results while keeping the deterministic viewpoint, lastly [78]. The last developments now tackle the still more involved problem of estimating not only the division rate but also the fragmentation kernel (i.e., how the sizes of the offspring are related to the size of the dividing individual) [13]. In parallel, in a long-run collaboration with statisticians, we studied the Piecewise Deterministic Markov Process (PDMP) underlying the equation, and estimated the division rate directly on sample observations of the process, thus making a bridge between the PDE and the PDMP approach in [95], a work which inspired also very recently other groups in statistics and probability [75], [105] and was the basis for Adélaïde Olivier's Ph.D thesis [122], [106] and of some of her more recent works [123] (see also axis 5).

**Data assimilation and stochastic modeling for protein aggregation**

Estimating reaction rates and size distributions of protein polymers is an important step for understanding the mechanisms of protein misfolding and aggregation (see also axis 5). In [63], we settled a framework problem when the experimental measurements consist in the time-dynamics of a moment of the population.

To model the intrinsic variability among experimental curves in aggregation kinetics - an important and poorly understood phenomenon - Sarah Eugène's Ph.D, co-supervised by P. Robert [99], was devoted to the stochastic modeling and analysis of protein aggregation, compared both with the deterministic approach traditionally developed in Mamba [137] and with experiments.

**Statistical methods decide on subsequently validated mechanism of ammonia detoxification**

To identify the mechanisms involved in ammonia detoxification [101], 8 candidate models representing the combination of three possible mechanisms were developed (axis 5). First, the ability of each model to capture the experimental data was assessed by statistically testing the null hypothesis that the data have been generated by the model, leading to exclusion of one of the 8 models. The 7 remaining models were compared among each other by the likelihood ratio. The by far best models were those containing a particular ammonia sink mechanism, later validated experimentally (axis 5). For each of the statistical tests, the corresponding test statistics has been calculated empirically and turned out to be not chi2-distributed in opposition to the usual assumption stressing the importance of calculating the empirical distribution, especially when some parameters are unidentifiable. This year the ammonia detoxification mechanisms have been integrated in a spatial-temporal agent-based model of a liver lobule (the smallest repetitive anatomical unit of liver) and studied for normal and fibrotic liver.

**Collaborations**

- **Marc Hoffmann,** Université Paris-Dauphine, for the statistical approach to growth and division processes [95], **M. Escobedo**, Bilbao and **M. Tournus**, Marseille, for the deterministic approach.
- **Tom Banks**, North Carolina State University, and **Philippe Moireau,** Inria M3DISIM, for the inverse problem and data assimilation aspects [70], [62]
- **Jan G. Hengstler**, IfADo, Dortmund, Germany

# 4. Application Domains

## 4.1. Introduction

The team has two main application-driven research axes.

*Applicative axis 1* focuses on cancer, an application on which almost all team members work, with various approaches. A main focus of the team is to study cancer as a Darwinian evolutionary phenomenon in phenotype-structured cell populations. Optimal control methods take into account the two main pitfalls of clinical cancer therapeutics, namely unwanted toxic side effects in healthy cell populations and drug resistance in cancer cell populations. Other studies concern telomere shortening, and multi-scale models.

*Applicative axis 2* is devoted to growth, evolution and regeneration in populations and tissues. It involves protein aggregation and fragmentation models for neurodegenerative diseases (prion, Alzheimer), organ modeling, mainly of the liver, its damages induced by toxic molecules, and its regeneration after toxic insult. Newcomers in this applicative field are epidemiological modeling of propagation of insect vector-borne diseases by reaction-diffusion equations and of their optimal control, bacterial growth and wound healing.

## 4.2. Applicative axis 1: Focus on cancer

**Personnel**

Luis Almeida, Cécile Carrère, Jean Clairambault, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard.

**Project-team positioning**

The MAMBA team designs and analyses mathematical models of tumor growth and therapy, at the cell population level, using agent-based or partial differential equations, with special interest in methodologies for therapeutic optimisation using combined anticancer drug treatments. Rather than, or not only, modeling the effect of drugs on molecular targets, we represent these effects by their *functional* consequences on the fate of healthy and cancer cell populations: proliferation (velocity of the cell division cycle, decreasing it, e.g., by antagonizing growth factor receptors), apoptosis, cell death or senescence.

Our goal in doing this is to circumvent the two main issues of anticancer therapy in the clinic, namely unwanted toxic side effects in populations of healthy cells and emergence of drug-induced drug resistance in cancer cell populations. This point of view leads us to take into account phenomena of transient and reversible resistance, observed in many cancer cell populations, by designing and analyzing models of cell populations structured in continuous phenotypes, relevant for the description of the behavior of cell populations exposed to drugs: either degree of resistance to a given drug, or potential of resistance to drug-induced stress, proliferation potential, and plasticity.

Such modeling options naturally lead us to to take into account in a continuous way (i.e., by continuous-valued phenotype or relevant gene expression) the wide phenotypic heterogeneity of cancer cell populations. They also lead us to adopt the point of view of *adaptive dynamics* according to which characteristic traits of cell populations evolve with tumor environmental pressure (drugs, cytokines or metabolic conditions, mechanical stress and spatial conditions), in particular from drug sensitivity to resistance. This position is original on the international scene of teams dealing with drug resistance in cancer.

**Scientific achievements**

**Modeling Acute Myeloid Leukemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations**
In collaboration with Catherine Bonnet (Inria DISCO, Saclay) and François Delhommeau (St Antoine hospital in Paris), together with DISCO PhD students José Luis Avila Alonso and Walid Djema, this theme has led to common published proceedings of conferences: IFAC, ACC, CDC, MTNS [66], [67], [68], [77], [91], [65]. These works study the stability of the haematopoietic system and its possible restabilization by combinations of anticancer drugs with functional targets on cell populations: proliferation, apoptosis, differentiation.

**Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control**
We tackle the problem to represent and inhibit - using optimal control algorithms, in collaboration with Emmanuel Trélat, proposed Inria team CAGE - drug-induced drug resistance in cancer cell populations. This theme, presently at the core of our works on cancer modeling with a evolutionary perspective on tumor heterogeneity, is documented in a series of articles [88], [89], [115], [116], [118]. Taking into account the two main pitfalls of cancer therapy, unwanted side effects on healthy cells and evolution towards resistance in cancer cells, it has attracted to our team the interest of several teams of biologists, with whom we have undertaken common collaborative works, funded by laureate answers to national calls (see ITMO Cancer HTE call).

This theme is also at the origin of methodological developments (see Research axis 1). In collaboration with Shensi Shen from Institut Gustave Roussy and Francois Vallette from Université de Nantes, we aim to develop simple non-spatial models to understand the mechanisms of drug resistance acquisition -and lost- in melanoma and glioblastoma. The models are systematically compared with in vitro and in vivo data generated by our collaborators and treated via image processing techniques developed in the team.

**Senescence modeling by telomere shortening**
In many animals, aging tissues accumulate senescent cells, a process which is beneficial to protect from cancer in the young organism. In collaboration with Teresa Teixeira and Zhou Xu from IBCP, we proposed a mathematical model based on the molecular mechanisms of telomere replication and shortening and fitted it on individual lineages of senescent Saccharomyces cerevisiae cells, in order to decipher the causes of heterogeneity in replicative senescence [79].

**Biomechanically mediated growth control of cancer cells**

Model simulations indicate that the response of growing cell populations on mechanical stress follows a simple universal functional relationship and is predictable over different cell lines and growth conditions despite the response curves look largely different. We developed a hybrid model strategy in which cells were represented by coarse-grained individual units calibrated in a high resolution cell model and parameterized each model cell by measurable biophysical and cell-biological parameters. Cell cycle progression in our model is controlled by volumetric strain, the latter being derived from a bio-mechanical relation between applied pressure and cell compressibility. After parameter calibration from experiments with mouse colon carcinoma cells growing against the resistance of an elastic alginate capsule, the model adequately predicts the growth curve in i) soft and rigid capsules, ii) in different experimental conditions where the mechanical stress is generated by osmosis via a high molecular weight dextran solution, and iii) for other cell types with different growth kinetics. Our model simulation results suggest that the growth response of cell population upon externally applied mechanical stress is the same, as it can be quantitatively predicted using the same growth progression function [44].

**Collaborations**

- AML modelling: **Catherine Bonnet**, DISCO Inria team, Saclay, and **François Delhommeau**, INSERM St Antoine (also collaborator in the INSERM HTE laureate project EcoAML, see below).
- INSERM HTE laureate project MoGlImaging, headed by E. Moyal (Toulouse): **François Vallette**, CRCNA and INSERM Nantes
- INSERM HTE laureate project EcoAML, headed by **François Delhommeau**, INSERM St Antoine: François Delhommeau, Thierry Jaffredo (IBPS), Delphine Salort (LCQB-IBPS)
- Adaptive dynamics to model drug resistance and optimal control to circumvent it:

  **Alexandre Escargueil, Michèle Sabbah** (1 PhD thesis in common), St Antoine Hospital, Paris

  **Emmanuel Trélat** (1 PhD thesis in common) at Inria team CAGE and Laboratoire Jacques-Louis Lions at Sorbonne Université.

  **Frédéric Thomas** at CREEC, Montpellier.

  **Tommaso Lorenzi** (Univ. of St Andrews).
- Telomere shortening: **Teresa Teixeira** and **Zhou Xu** (IBCP, Paris), **Philippe Robert** (Inria RAP).
- TRAIL treatment: **Gregory Batt,** Inria Saclay and Inst. Pasteur (France)
- Biomechanical control of cancer cells: **Pierre Nassoy**, Bioimaging and Optofluidics Group, LP2N – UMR 5298. IOGS, CNRS & University of Bordeaux

## 4.3. Applicative axis 2: Growth, evolution and regeneration in populations and tissues

**Personnel**

Luis Almeida, Pierre-Alexandre Bliman, Marie Doumic, Dirk Drasdo, Benoît Perthame, Diane Peurichard, Nastassia Pouradier Duteil, Philippe Robert

**Project-team positioning**

The applications in this category span very different subjects from amyloid diseases, dengue fever, wound healing, liver regeneration and toxicity, up to bacterial growth and development of organisms. As the applications, the methods span a wide range. Those concerning identification of models and parameters with regard to data have partially been outlined in axis 3. Focus in this axis is on the model contribution to the biologically and/or medically relevant insights and aspects.
Liver-related modeling is partially performed within the Inria team MIMESIS (Strasbourg) with the focus on real-time, patient-specific biomechanical liver models to guide surgery and surgeons. Internationally, spatial temporal liver related models are developed in Fraunhofer MEVIS (Bremen), by T. Ricken (University of Stuttgart), and P. Segers group (Leuven). Different from these, Mamba has a strong focus on spatial-temporal modeling on the histological scale, integration of molecular processes in each individual cell, and single-cell (agent) based models [32], [30], [143]. Works by Schliess [139], [101] have been highlighted in editorials.

Mathematical modeling of protein aggregation is a relatively recent domain, only a few other groups have emerged yet; among them we can cite the Inria team Dracula, with whom we are in close contact, and e.g., the work by Jean-Michel Coron (Sorbonne Université) and Monique Chyba (Hawaii, USA) in control, and Suzanne Sindi (USA) for the modeling of the yeast prion. We have interactions with all these groups and organized a workshop in June 2017, gathering both the biophysics and applied mathematics communities.

**Scientific achievements**

**Amyloid disease**
Application to protein aggregation in amyloid diseases is a long-standing interest of Mamba, dating back to 2010 [85], and developed through the collaboration with Human Rezaei's team at Inra. More recently, with Wei-Feng Xue in Canterbury, we investigated the intrinsic variability among identical experiments of nucleation [93], [100], Sarah Eugène's Ph.D subject (co-supervised by Philippe Robert) [99].

In collaboration with Tom Banks first [69], [70] and then Philippe Moireau, we developed quantitative comparisons between model and data. Through data assimilation and statistical methods [63], we proposed new models and mechanisms.

**Biological control of arboviroses**
Sterile Insect Technique (SIT) [98] is a biological control method relying on massive releases of sterile male insects into the wild. The latter compete with wild males to mate with the females, and induce no offspring to the latter, thus reducing the next generation's population. This can result in a progressive reduction, or even disparition, of the target population.

A related technique is based on the infection by *Wolbachia* [104]. This symbiotic bacterium is maternally transmitted from infected females to their offspring, but induces *cytoplasmic incompatibility* [141], [80]: mating between infected males and uninfected females gives no offspring. Releases of *Wolbachia* infected males alone is thus comparable to classical SIT.

On the other hand, releasing both infected males and females in sufficient quantity may result in infection of the wild population. This gives rise to an interesting new control principle, as *Wolbachia* has been shown to severely reduce the insect vectorial ability to transmit dengue, zika or chikungunya, indirectly by lifespan and fertility reduction, and directly by reducing the ability of the viruses to proliferate within the organism [121].

We proposed new insights on the practical and theoretical issues raised by the implementation of the previous methods.

**Wound healing 1: epithelial tissues**

We studied cell motion in epithelial gap closure, a form of collective cell migration that is a very widespread phenomenon both during development and adult life - it is essential for both the formation and for the maintenance of epithelial layers. Due to their importance, *in vivo* wound healing and morphogenetic movements involving closure of holes in epithelia have been the object of many studies. In our works [0] we considered wound healing and epithelial gap closure in both in vivo (in particular drosophila pupa) and in vitro (MDCK cell and human keratinocytes). We found some similarities in the geometry dependence of the wound closure strategies between these two situations, indicating the existence of conserved mechanisms that should be widespread across living beings. We are concentrating on the study of actin cable formation.

**Wound healing 2: adipose tissues**
After injury, if regeneration can be observed in hydra, planaria and some vertebrates, regeneration is rare in mammals and particularly in humans. In this research axis, we investigated the mechanisms by which biological tissues recover after injury. We explored this question on adipose tissue, using the mathematical framework recently developed in [134]. Our assumption is that simple mechanical cues between the Extra-Cellular Matrix (ECM) and differentiated cells can explain adipose tissue morphogenesis and that regeneration requires after injury the same mechanisms. We validated this hypothesis by means of a two-dimensional Individual Based Model (IBM) of interacting adipocytes and ECM fiber elements [135]. The model successfully generated regeneration or scar formation as functions of few key parameters, and seemed to indicate that the fate of injury outcome could be mainly due to ECM rigidity.

**Modeling of morphogen diffusion in *Drosophila* oogenesis**
In collaboration with a team of developmental biologists of Rutgers University (Camden, New Jersey), we have built a model for the diffusion of the Gurken morphogen during *Drosophila* oogenesis, taking into account a wide variety of biological mechanisms such as diffusion of the morphogen, reactions of components of the EGFR signaling pathway, movement of the source of morphogen, shift of the overlying follicle cells and growth of the egg chamber. This model, together with a complete numerical code developed in Matlab, provides a tool to understand how each mechanism influences the signal distribution. The overall aim of the project is to use this tool to guide future experiments, and to understand what mechanisms contribute to the different distributions of signal among species.

**Bacterial population growth**
We exploited all the methods developed to estimate the division rate of a population (see axis 3) to address a seminal question of biology: is it a size-sensing or a timing mechanism which triggers bacterial growth? In [138], we showed that a sizer model is robust and fits the data well. Several studies from other groups came at the same time, showing a renewed interest on a question dated back to Jacques Monod's PhD thesis (1941). Of special interest is the "adder" model, for which we are currently developing new estimation methods.

**A quantitative high resolution computational mechanics cell model for growing and regenerating tissues**
Mathematical models are increasingly designed to guide experiments in biology, biotechnology, as well as to assist in medical decision-making. They are in particular important to understand emergent collective cell behavior. For this purpose, the models, despite still abstractions of reality, need to be quantitative in all aspects relevant for the question of interest. During the regeneration of liver after drug-induced depletion of hepatocytes surviving dividing and migrating hepatocytes must squeeze through a blood vessel network to fill the emerged lesions. Here, the cells' response to mechanical stress might significantly impact on the regeneration process. We developed a 3D high-resolution cell-based model integrating information from measurements in order to obtain a refined quantitative understanding of the cell-biomechanical impact on the closure of drug-induced lesions in liver. Our model represents each cell individually, constructed as a physically scalable network of viscoelastic elements, capable of mimicking realistic cell deformation and supplying information at subcellular scales. The cells have the capability to migrate, grow and divide, and infer the nature of their mechanical elements and their parameters from comparisons with optical stretcher experiments. Due to triangulation of the cell surface, interactions of cells with arbitrarily shaped (triangulated) structures such as blood vessels can be captured naturally. Comparing our simulations with those of so-called center-based models, in which cells have a rigid shape and forces are exerted between cell centers, we find

[0] ravasio:hal-01245750, vedula:hal-01298859

that the migration forces a cell needs to exert on its environment to close a tissue lesion, is much smaller than predicted by center-based models. This effect is expected to be even more present in chronic liver disease, where tissue stiffens and excess collagen narrows pores for cells to squeeze through [44].

Main collaborators: Stefan Höhme, Univ. Leipzig; Josef Käs, Univ. Leipzig.

**Modeling the extracellular matrix in multicellular organization and liver regeneration:**
An important step has been undertaken to integrate an explicit model of collagen networks in liver and other tissues. The mechanical model of collagen fibers uses linear and rotational springs to represent collagen fibers and collagen network taking into account the stretching and bending energy of the collagen network. The model has been validated by direct comparison to experiments where a force has been excerted on a single collagen fiber, as well as to shear experiments. In a next step, this collagen model has been incorporated into the previous lobule model to simulate spatio-temporal pattern of the ECM in the lobule. One key objective is a model of fibrosis development, whereby fibrotic streets form. So far, not coherent model exist but a number of hypotheses that will be implemented and tested versus data.

Main collaborators: Steven Dooley, Seddik Hammad, Univ. Mannheim; JG Hengstler, IfADo.

**Models of flow in liver**
Also for the liver, a model for bile salts transport has been developed. The current hypothesis is that bile salt excreted by hepatocytes into bile canaliculi, are transported within canaliculi by convection through the canal of hering to the bilary ducts. In close iterations with experiments and by image based modeling runing simulations directly in reconstructed 3D volume data sets, we test different alternative mechanisms of bile transport.

The blood flow model of the individual liver lobule, the smallest anatomical and functional repetitive unit of liver has been embedded in an electrical analogue model for the whole model hemodynamics to compute the impact of architectural changes at the lobule level as they occur after partial hepatectomy on the whole body hemodynamics. At the lobule level, the impact of capillary diameters on the flow have been studied under consideration of the hematocrit value (the volume fraction of red blood cells).

Collaborators: Chloé Aubert, UPMC, I. Vignon-Clementel, REO, Jan G. Henstler and Natiket Vartak, IfADo, Eric Vibert, Hopital Paul Brousse, Villejuif

**Liver development**
The deformable cell model is being used to establish a model of bile duct formation. Bile ducts from at the portal veins as a consequence of an interaction of cholagiocytes, aligning the mesenchyme of the portal vein, and hepatoblasts surrounding the portal veins. Hepatoblasts are a pre-stage of hepatocytes. Current biological hypotheses speculate that bile ducts may either emerge from proliferation of the hepatoblast layer that is in contact to cholangiocytes, leading to formation of a lumen by buckling, by attraction of water a positions where cholangiocytes secret mucin, or by apical contraction of hepatoblasts within the layer in contact to the cholangiocytes. The simulations are performed with the deformable cell model.

Collaborators: Frederic Lemaigre, De Duve Institut, Brussels, Anne Dubart-Kupperschmitt, Hopital Paul Brousse.

**Image processing, analysis and quantification of tissue microarchitecture**
At the interface between experiment and modeling we pursue a number of projects on image analysis and quantification. Such information in the past often served to generate hypotheses of the mechanisms underlying image sequences in time, which then could be turned in a mathematical model to verify, which hypotheses are sufficient to explain the image data. Several image analysis projects focus on the liver. (1.) Bile microinfacts have been found to be initiated by rupture of the apical hepatocyte membrane leading to shunting from bile canaliculi to the blood capillaries. This is followed by massive increase of the immune cells as could be quantified by analysis of intravital micrographs. For every frame in the video, a binary mask that most likely resemble detected immune cells were obtained. The process started by applying suitable linear and non-linear filters to highlight structure of interest and remove noise. Morphological operations and blob analysis was finally utilized to locate and count the cells. With the help of an expert, the confusion matrix was finally established to assess the quality of the segmentation and the obtained results (Ghallab et. al., J. Hepat.

2018; https://aasldpubs.onlinelibrary.wiley.com/doi/full/10.1002/hep.30213. [14]) (2.) In a second project, the CYP – enzyme distribution is quantified after repetitive administration of CCl4. After single overdose of CCl4, the liver shows a peri-central liver lobule (smallest repetitive anatomical unit of liver) necrosis. This pattern changes in repetitive dosing and leads to chronic disease stages and sometimes eventually to either hepatocellular carcinoma [15] or acute-on-chronic liver failure (ACLF), a disease condition with often-lethal outcome. Data from whole-slide scans is analyzed to serve to develop a mathematical model of ACLF. (3.) A similar strategy is pursued for fibrosis formation through high fat diet both by image analysis of mouse and human data, aiming at a mathematical lobule model based on a deformable cell model. Here currently images are analyzed to quantify microarchitectural modifications as a consequence of Western Diet (a high fat diet generating a disease condition reminiscent of NAFLD in human. (4.) TiQuant-algorithms have been used to analyze micro-and macrovascular alterations in cirrhosis [125] and (5.) tissue modifications after PHx [82].

Main collaborators: Ahmed Ghallab, Jan G. Hengstler, IfADo; Ursula Klingmüller, DKFZ Heidelberg; Steven Dooley, Univ. Hospital Mannheim; Percy Knolle, Helmholz Inst. Munich, Joachim Bode, Univ. Hospital Düsseldorf, Christian Trautwein, Univ. Hosp. Aachen; P. Seegers (Ghent University); Eric Vibert (Hopital Paul Brousse).

**Relating imaging on microscopic scales with imaging on macroscopic scales: From Diffusion-Weighted MRI Calibrated With Histological Data: an Example From Lung Cancer**

Diffusion-weighted magnetic resonance imaging (DWI) is a key non-invasive imaging technique for cancer diagnosis and tumor treatment assessment, reflecting Brownian movement of water molecules in tissues. Since densely packed cells restrict molecule mobility, tumor tissues produce usually higher signal (less attenuated signal) on isotropic maps compared with normal tissues. However, no general quantitative relation between DWI data and the cell density has been established. In order to link low-resolution clinical cross-sectional data with high resolution histological information, we developed an image processing and analysis chain, which was used to study the correlation between the diffusion coefficient (D value) estimated from DWI and tumor cellularity from serial histological slides of a resected non-small cell lung cancer tumor. Color deconvolution followed by cell nuclei segmentation was performed on digitized histological images to determine local and cell-type specific 2d (two-dimensional) densities. From these, the 3d cell density was inferred by a model-based sampling technique, which is necessary for the calculation of local and global 3d tumor cell count. Next, DWI sequence information was overlaid with high resolution CT data and the resected histology using prominent anatomical hallmarks for co-registration of histology tissue blocks and non-invasive imaging modalities' data. The integration of cell numbers information and DWI data derived from different tumor areas revealed a clear negative correlation between cell density and D value. Importantly, spatial tumor cell density can be calculated based on DWI data. In summary, our results demonstrate that tumor cell count and heterogeneity can be predicted from DWI data, which may open new opportunities for personalized diagnosis and therapy optimization [145]. The work of that paper has been further advanced to adapt the procedures for clinical use (in preparation).

**Collaborations**

- Biological control of arboviroses: **Nicolas Vauchelet** (Université Paris 13); **Grégoire Nadin** (LJLL, Sorbonne Université); **Yannick Privat** (Université de Strasbourg); **D. Villela, C. Struchiner** (Fiocruz, Brazil); **Jorge Zubelli** (IMPA, Brazil); **Alain Rapaport** (INRA-Montpellier), **Y. Dumont** (CIRAD-Montpellier); **Ch. Schaerer, P. Pérez-Estigarribia** (UNA, Paraguay), **O. Vasilieva** (Universidad del Valle, Cali, Colombia), **D. Cardona-Salgado** (Universidad Autónoma de Occidente, Cali, Colombia).
- Protein aggregation in amyloid diseases: **Human Rezaei**'s team at Inra Jouy-en-Josas (France) and **W-F Xue**'s team in at university of Kent (Great Britain); **Tom Banks** at the North Carolina State University (USA) and **Philippe Moireau** (M3DISIM)
- Bacterial growth and division: **Lydia Robert**, Sorbonne Université (France)
- Liver research & toxicology: **JG. Hengstler** group (IfADo, Dortmund, Germany); **R. Gebhardt** (Univ. Leipzig); **U. Klingmueller** (DKFZ, Heidelberg); **Irène Vignon-Clementel** (Inria, REO)

- Wound healing: **Patrizia Bagnerini** (Genova, Numerical methods), **Benoît Ladoux** (Institut Jacques Monod et Mechanobiology Institute Singapore, Biophysics) and **Antonio Jacinto** (CEDOC, Lisbon, Biology and Medicine). (Adipose tissue regeneration) team of **L. Casteilla** (StromaLab, Toulouse)
- Diffusion of morphogen: Center for Computational and Integrative Biology, Rutgers University (Camden, New Jersey), joint work with Professor Nir Yakoby's Drosophila Laboratory
- Linking micro and macro-image information: Oliver Sedlaczek, Univ. and DKFZ Heidelberg, Kai Breuhahn, Univ. Heidelberg.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

We welcome a new team member, Nastassia Pouradier-Duteil, junior research scientist since September 2018.

We welcome Ayman Moussa in delegation since September 2018; he defended his habilitation thesis on December 13th.

Marie Doumic finished her two-year sabbatical stay in September 2018.

Jean Clairambault is emeritus DR since March 2018.

### *5.1.1. Awards*

In December 5, 2017, Benoit Perthame has been elected at the Académie des Sciences, and was received in the Académie on May 28, 2018.

Christian Schmeiser, associate member of Mamba through the associated team MaMoCeMa with the university of Vienna, being the laureate of the "chaire d'excellence" of the FSMP, is for six months in Paris (september 2018 to february 2019).

# 6. New Software and Platforms

## 6.1. TiQuant

*Tissue Quantifier*
KEYWORDS: Systems Biology - Bioinformatics - Biology - Physiology
FUNCTIONAL DESCRIPTION: Systems biology and medicine on histological scales require quantification of images from histological image modalities such as confocal laser scanning or bright field microscopy. The latter can be used to calibrate the initial state of a mathematical model, and to evaluate its explanatory value, which hitherto has been little recognized. We generated a software for image analysis of histological material and demonstrated its use in analysing liver confocal micrografts, called TiQuant (Tissue Quantifier). The software is part of an analysis chain detailing protocols of imaging, image processing and analysis in liver tissue, permitting 3D reconstructions of liver lobules down to a resolution of less than a micrometer.

- Author: Dirk Drasdo
- Contact: Dirk Drasdo

## 6.2. TiSim

*Tissue Simulator*
KEYWORDS: Systems Biology - Bioinformatics - Biology - Physiology

SCIENTIFIC DESCRIPTION: TiSim (Tissue Simulator) is a versatile and efficient simulation environment for tissue models. TiSim is a software for agent-based models of multicellular systems. It permits model development with center-based models and deformable cell models, it contains modules for monolayer and multicellular spheroid simulations as well as for simulations of liver lobules. Besides agent-based simulations, the flow of blood and the transport of molecules can be modelled in the extracellular space, intracellular processes such as signal transduction and metabolism can be simulated, for example over an interface permitting integration of SBML-formulated ODE models. TiSim is written in modern C++ , keeping central model constituents in modules to be able to reuse them as building blocks for new models. For user interaction, the GUI Framework Qt is used in combination with OpenGL for visualisation. The simulation code is in the process of being published. The modeling strategy and approaches slowly reach systems medicine and toxicology. The diffusion of software is a fundamental component as it provides the models that are complex and difficult to implement (implementing a liver lobule model from scratch takes about 2-2.5yrs) in form of a software to the developer and users who like to build upon them. This increases significantly the speed of implementing new models. Moreover, standardization is indispensible as it permits coupling different software tools that may have implemented models at different scales / levels.
FUNCTIONAL DESCRIPTION: TiSim is a software that permits agent-based simulations of multicellular systems. - center-based lattice-free agent-based model - modular - C++, Qt, OpenGL, GUI, batch mode - permits multiscale simulations by integration of molecular pathways (for signaling, metabolisms, drug) into each individual cell - applications so far: monolayer growth, multicellular spheroids - Boolean networks (development time = coding time ( 60 MMs) + model development time ( 264 MMs)) - in follow-up version 1: - liver lobule regeneration - SBML interface - in follow-up version 2: - deformable cell model (by triangulation of cell surface) - deformable rod models - extracellular matrix - vascular flow and transport TiSim can be directly fed by processed image data from TiQuant.

- Participants: Andreas Buttenschoen, Dirk Drasdo, Eugenio Lella, Géraldine Cellière, Johannes Neitsch, Margaretha Palm, Nick Jagiella, Noémie Boissier, Paul Van Liedekerke, Stefan Hoehme and Tim Johann
- Partner: IZBI, Université de Leipzig
- Contact: Dirk Drasdo

## 6.3. Platforms

### *6.3.1. TiSim*

New side branches have been developed that integrate the deformable cell model and extracellular matrix model in the TiSim software.

### *6.3.2. TiQuant*

The software has been further extended for machine learning components.

# 7. New Results

## 7.1. Modelling Polymerization Processes

**Nucleation Phenomena.**
A new stochastic model of polymerization including the nucleation has been analyzed in [4]. A Functional Central Limit Theorem for the Becker-Döring model in an infinite dimensional state space is established in [25].

**An oscillatory model of polymerisation-depolymerisation.**

In 2017, we evidenced the presence of several polymeric species by using data assimilation methods to fit experimental data from H. Rezaei's lab [64]. In collaboration with Klemens Fellner from the university of Graz, we now propose a new model, variant of the Becker-Döring system but containing two monomeric species, capable of displaying sustained though damped oscillations [39].

**Time asymptotics for nucleation, growth and division equations.**
We revisited the well-known Lifshitz-Slyozov model, which takes into account only polymerisation and depolymerisation, and progressively enriched the model. Taking into account depolymerisation and fragmentation reaction term may surprisingly stabilisde the system, since a steady size-distribution of polymers may then emerge, so that "Ostwald ripening" does not happen [8].

**Cell population dynamics and its control**
The PhD thesis work of Camille Pouchol (co-supervisors Jean Clairambault, Michèle Sabbah, INSERM, and Emmanuel Trélat, Inria CAGE and LJLL) has been continued, leading after his first article published in the J. Maths Pures Appl. [136], summarised in [31], to his PhD defence in June [1], and to a diversification of his research activities in various directions related to population dynamics and optimal control with Antoine Olivier, Emmanuel Trélat and Enrique Zuazua [51], [56] or to more general questions [55].

**Measure solutions for the growth-fragmentation equation**
As recalled in the section "Foundations", entropy methods for population dynamics have been successfully developed around B. Perthame and co-authors. We recently extend such methods to the growth-fragmentation equation, in collaboration with P. Gwiazda, E. Wiedemann and T. Debiec [40], using the framework of generalised Young measures.

## 7.2. Large Stochastic Networks

The equilibrium properties of allocation algorithms for networks with a large number of nodes with finite capacity are investigated in [46] and in [60].

## 7.3. Control Strategies for Sterile Insect Techniques

We proposed different models to serve as a basis for the design of control strategies relying on releases of sterile male mosquitoes (*Aedes spp*) and aiming at elimination of wild vector population. Different types of releases were considered (constant, periodic or impulsive) and sufficient conditions to reach elimination were provided in each case [57], [3], [35]. We also estimated sufficient and minimal treatment times. A feedback approach was introduced, in which the impulse amplitude is chosen as a function of the actual wild population [57], [3], [35].

## 7.4. Optimal replacement strategies, application to Wolbachia

We modelled and designed optimal release control strategy with the help of a least square problem. In a nutshell, one wants to minimize the number of uninfected mosquitoes at a given time horizon, under relevant biological constraints. We derived properties of optimal controls and studied a limit problem providing useful asymptotic properties of optimal controls [49], [3].

## 7.5. Oscillatory regimes in population models

Understanding mosquitoes life cycle is of great interest presently because of the increasing impact of vector borne diseases. Observations yields evidence of oscillations in these populations independent of seasonality, still unexplained. We proposed [58], [3] a simple mathematical model of egg hatching enhancement by larvae which produces such oscillations that conveys a possible explanation.

On the other hand, population oscillations may be induced by seasonal changes. We considered a biological population whose environment varies periodically in time, exhibiting two very different "seasons", favorable and unfavorable. We addressed the following question: the system's period being fixed, under what conditions does there exist a critical duration above which the population cannot sustain and extincts, and below which the system converges to a unique periodic and positive solution? We obtained [59], [3] sufficient conditions for such a property to occur for monotone differential models with concave nonlinearities, and applied the obtained criterion to a two-dimensional model featuring juvenile and adult insect populations.

## 7.6. Feedback control principles for population replacement by Wolbachia

The issue of effective scheduling of the releases of *Wolbachia*-infected mosquitoes is an interesting problem for Control theory. Having in mind the important uncertainties present in the dynamics of the two populations in interaction, we attempted to identify general ideas for building release strategies, which should apply to several models and situations [34]. These principles were exemplified by two interval observer-based feedback control laws whose stabilizing properties were demonstrated when applied to a model retrieved from [76].

## 7.7. Bacterial motion by run and tumble

Collective motion of chemotactic bacteria such as Escherichia coli relies, at the individual level, on a continuous reorientation by runs and tumbles. It has been established that the length of run is decided by a stiff response to a temporal sensing of chemical cues along the pathway. We describe in [21] a novel mechanism for pattern formation stemming from the stiffness of chemotactic response relying on a kinetic chemotaxis model which includes a recently discovered formalism for the bacterial chemotaxis. We prove instability both for a microscopic description in the space-velocity space and for the macroscopic equation, a flux-limited Keller-Segel equation, which has attracted much attention recently. A remarkable property is that the unstable frequencies remain bounded, as it is the case in Turing instability. Numerical illustrations based on a powerful Monte Carlo method show that the stationary homogeneous state of population density is destabilized and periodic patterns are generated in realistic ranges of parameters. These theoretical developments are in accordance with several biological observations.

This motivates also our study of traveling wave and aggregation in population dynamics of chemotactic cells based on the FLKS model with a population growth term [7]. Our study includes both numerical and theoretical contributions. In the numerical part, we uncover a variety of solution types in the one-dimensional FLKS model additionally to standard Fisher/KPP type traveling wave. The remarkable result is a counter-intuitive backward traveling wave, where the population density initially saturated in a stable state transits toward an un- stable state in the local population dynamics. Unexpectedly, we also find that the backward traveling wave solution transits to a localized spiky solution as increasing the stiffness of chemotactic response. In the theoretical part, we obtain a novel analytic formula for the minimum traveling speed which includes the counter-balancing effect of chemotactic drift vs. reproduction/diffusion in the propagating front. The front propagation speeds of numerical results only slightly deviate from the minimum traveling speeds, except for the localized spiky solutions, even for the backward traveling waves. We also discover an analytic solution of unimodal traveling wave in the large-stiffness limit, which is certainly unstable but exists in a certain range of parameters.

## 7.8. Numerical methods for cell aggregation by chemotaxis

Three-dimensional cultures of cells are gaining popularity as an in vitro improvement over 2D Petri dishes. In many such experiments, cells have been found to organize in aggregates. We present new results of three-dimensional in vitro cultures of breast cancer cells exhibiting patterns. Understanding their formation is of particular interest in the context of cancer since metastases have been shown to be created by cells moving in clusters. In the paper [37], we propose that the main mechanism which leads to the emergence of patterns is chemotaxis, i.e., oriented movement of cells towards high concentration zones of a signal emitted by the cells themselves. Studying a Keller-Segel PDE system to model chemotactical auto-organization of cells, we prove that it is subject to Turing instability if a time-dependent condition holds. This result is illustrated by two-dimensional simulations of the model showing spheroidal patterns. They are qualitatively compared to the biological results and their variability is discussed both theoretically and numerically.

This motivates to study parabolic-elliptic Keller-Segel equation with sensitivity saturation, because of its pattern formation ability, is a challenge for numerical simulations. We provide in [16] two finite-volume schemes that are shown to preserve, at the discrete level, the fundamental properties of the solutions, namely energy dissipation, steady states, positivity and conservation of total mass. These requirements happen to be critical when it comes to distinguishing between discrete steady states, Turing unstable transient states, numerical artifacts or approximate steady states as obtained by a simple upwind approach. These schemes are obtained either by following closely the gradient flow structure or by a proper exponential rewriting inspired by the Scharfetter-Gummel discretization. An interesting fact is that upwind is also necessary for all the expected properties to be preserved at the semi-discrete level. These schemes are extended to the fully discrete level and this leads us to tune precisely the terms according to explicit or implicit discretizations. Using some appropriate monotonicity properties (reminiscent of the maximum principle), we prove well-posedness for the scheme as well as all the other requirements. Numerical implementations and simulations illustrate the respective advantages of the three methods we compare.

## 7.9. Focus on cancer

**Modelling Acute Myeloid Leukaemia (AML) and its control by anticancer drugs by PDEs and Delay Differential equations**

This theme has continued to be developed in collaboration with Catherine Bonnet, Inria DISCO (Saclay) [12], [29]. Without control by drugs, but with representation of mutualistic interactions between tumor cells and their surrounding support stroll cells, it has also, in collaboration with Delphine Salort and Thierry Jaffredo (LCQB-IBPS) given rise to a recent work by Thanh Nam Nguyen, hired as HTE and ERC postdoctoral fellow at LCQB, submitted as full article [50].

**Adaptive dynamics setting to model and circumvent evolution towards drug resistance in cancer by optimal control**

The research topic “Evolution and cancer”, designed in the framework of adaptive dynamics to represent and overcome acquired drug resistance in cancer, initiated in [119], [118] and later continued in [90], [89], [117], has been recently summarised in [31] and has been the object of the PhD thesis work of Camille Pouchol, see above “Cell population dynamics and its control” . It is now oriented, thanks to work underway by Cécile Carrère, Jean Clairambault, Tommaso Lorenzi and Grégoire Nadin, in particular towards the mathematical representation of *bet hedging* in cancer, namely a supposed optimal strategy consisting for cancer cell populations under life-threatening cell stress in diversifying their phenotypes according to several resistance mechanisms, such as overexpression of ABC transporters (P-glycoprotein and many others), of DNA repair enzymes or of intracellular detoxication processes. According to different deadly insults the cancer cell population is exposed to, some phenotypes may be selected, any such successful subpopulation being able to store the cell population genome (or subclones of it if the cell population is already genetically heterogeneous) and make it amenable to survival and renewed replication.

**Philosophy of cancer biology**

This new research topic in Mamba, dedicated to explore possibly underinvestigated, from the mathematical modelling point of view, parts of the field of cancer growth, evolution and therapy, has been the object of a presentation by Jean Clairambault at the recent workshop “Philosophy of cancer biology’ (https://www.philinbiomed.org/event/philosophy-of-cancer-biology-workshop/). This workshop gathered most members worldwide of this small, but very active in publishing, community of philosophers of science whose field of research is “philosophy of cancer”, as they call it themselves. This topic offers a clear point of convergence between mathematics, biology and social and human sciences.

## 7.10. Deformable Cell Modeling: biomechanics and Liver regeneration

- Biomechanically mediated growth control of cancer cells The key intriguing novelty was that the same agent-based model after a single parameter has been calibrated with growth data for multicellular spheroids without application of external mechanical stress by adapting a single parameter, permitted to correctly predict the growth speed of multicellular spheroids of 5 different cell lines subject of external mechanical stress. Hereby the same mechanical growth control stress function was used without any modification [44]. The prediction turned out to be correct independent of the experimental method used to exert the stress, whereby once a mechanical capsule has been used, once dextran has been used in the experiments.
- Regeneration of liver with the Deformable Cell Model. The key novelty was the implementation of the model itself, but an interesting novel result is that the DCM permits closure of a pericentral liver lobule lesion generated by drug-induced damage with about 5 times smaller active migration force due to the ability of the cell to strongly deform and squeeze into narrow spaces between the capillaries. This finding stresses that a precise mechanical description is important in view of quantitatively correct modeling results [142]. The deformable cell model however could be used to calibrate the interaction forces of the computationally much cheaper center-based model to arrive at almost the same results.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

Contract with Orange labs (2016-2018) for Veronica Quintuna's PhD. See Reference [2].

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR*

#### *9.1.1.1. ANR Blanc 2014-2018 "Kibord"*

This project gathers several members of the MAMBA team together with the ENS Cachan and Université Paris-Dauphine on the mathematical study of PDE models with application to biology.

#### *9.1.1.2. ANR iLITE 2016 - 2020*

Jean-Charles Duclos-Vallée, Paul Brousse Hospital, Villejuif. Partners are several departments in Paul Brousse Hospital, ENS Cachan, University of Compiègne and several companies all over France, and REO team, Inria Paris. The pursued objective is the bioengineering design of an artificial liver intended for liver replacement.

#### *9.1.1.3. ANR InTelo 2017-2020*

Telomere dynamics, headed by Teresa Teixeira (IBPC, Paris).

#### *9.1.1.4. INCa/DGOS; PRT-K 2018-2021*

Khê HOANG-XUAN, Hôpital Universitaire La Pitié Salpêtrière, Paris. Mathematical modeling at micro and macroscopic level of primary central nervous system lymphomas (PCNSL).

### *9.1.2. ITMO Cancer 2016 - 2020, HTE call (heterogeneity of tumours in their ecosystems)*

#### *9.1.2.1. ITMO Cancer EcoAML*

Early leukaemogenesis in Acute Myelogenous Leukaemia (AML), 8 teams headed by François Delhommeau (CDR St Antoine, Paris).

*9.1.2.2. ITMO Cancer MoGlImaging*

Treatment-induced treatment resistance and heterogeneity in glioblastoma, 8 teams headed by Elizabeth Moyal (INSERM, Toulouse).

# 9.2. European Initiatives

## *9.2.1. Collaborations in European Programs, Except FP7 & H2020*

Program: Celtic+

Project acronym: Sendate

Project title:Secure Networking for a Data Center Cloud in Europe

April 2016/May 2019

Coordinator: Nokia

Other partners: Siemens, IMT, ...

# 9.3. International Initiatives

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

*9.3.1.1. MaMoCeMa*

Title: Mathematical modeling of cell motility and of autophagy

International Partner (Institution - Laboratory - Researcher):

University of Vienna (Austria) - Wolfgang Pauli Institute - Christian Schmeiser

Start year: 2018

Numerous fruitful collaborations have been developed these last years between the WPI and the Inria team MAMBA. Diane Peurichard – newly recruited permanent member of the team MAMBA – worked two years (2016-2017) with Christian Schmeiser – member of the present project – through a post-doctoral contract at the university of Vienna. In collaboration with the biologists of IST, they developed mathematical tools to understand how cells move through adhesion-based and adhesion-free motion with applications in cancer development, prevalent theme of the team MAMBA. Collaborations WPI-MAMBA are presently maintained and ensured by the sabbatical of Marie Doumic – MAMBA team leader –, working at the university of Vienna with Christian Schmeiser and the PhD student Julia Delacour. They have initiated a collaboration on the mathematical modeling of autophagy, which requires both C. Schmeiser's expertise in biomechanics and M. Doumic's knowledge on aggregation processes. This team will also benefit of the strong links that C. Schmeiser has developed with the two biologists teams of S. Martens (on autophagy) and M. Sixt (on cell movement).

Of note, C. Schmeiser has been a laureate for the "Chaire d'excellence" program of the FSMP. As such, he is for six months in Paris, and delivered a course at IHP on entropy methods. Many of his students and collaborators visited him (D. Oelz, G. Jankowiak, L. Kanzler, G. Favre, L. Neumann...), and participated to a joint Mamba-MaMoCeMa meeting on December 6th, still strengthening our links.

## *9.3.2. Participation in Other International Programs*

*9.3.2.1. International Initiatives*

**ECOS Nord C17M01**

Title: News methods for controle of dengue and arobivroses epidemics

International Partner (Institution - Laboratory - Researcher):

Universidad del Valle (Colombia) - Department of Mathematics - Olga Vasilieva

Duration: 2017 - 2019

Start year: 2017

The overall goal of the project is the development of mathematical models and theory-based control methods, contributing to the improvement of epidemiological surveillance and the control of dengue and other serious diseases transmitted by mosquitoes Aedes aegypti (chikungunya, yellow fever, zika fever). More specifically, it :

- Develops modeling framework for the biological control of mosquito populations (through the use of natural predators, Wolbachia bacteria etc.).
- Proposes and evaluates control strategies based on the use of biological agents and on their possible combinations with traditional control measures (such as removal of reproduction, spraying insecticides and / or larvicides, use of mosquito nets, repellents, etc.).
- Compares the results of biological control strategies (and their combinations) with those of traditional control using a cost / efficiency approach.
- Includes in the developments the spatial aspects of the questions above.

**BMBF (Germany) / LiSym; 2016-2020** LiSym addresses liver diseases and regeneration, namely, steatosis, fibrosis and cirrhosis, and acute on chronic liver failure. Dirk Drasdo is co-coordinator of one sub-project, participant in one of the other ones, and member of the leadership board

**BMBF (Germany) / MSDILI; 2016-2019** MS-DILI addresses multiscale modeling of drug-induced liver disease focusing on the role of APAP. Dirk Drasdo participates in this project.

## 9.4. International Research Visitors

### *9.4.1. Visits of International Scientists*

- Prof. Olga Vasilieva (Universidade del Valle, Cali, Colombia) was invited during three weeks, together with Edwin Bairros, PhD student.
- Prof. Yukihiko Nakata (Shimane University, Matsue, Japan) was hosted during one week in the framework of the French program Exploration France.
- Prof. C. Schmeiser (university of Vienna, Austria) was visiting during four month, from september 2018, and should stay until february 2019.
- Prof. D. Oelz (university of Queensland, Australia) visited from Dec. 5th to Dec. 21st.
- Jieling Zhao, Postdoc from IfADo
- Paul van Liedekerke, Research engineer from IfADo

*9.4.1.1. Internships*

Ismael Gonzalez Valverde (University of Zaragoza) visited our team for 3 months working on implementation of the meshing of liver micro-structures in modeling of liver regeneration within TiSim.

### *9.4.2. Visits to International Teams*

*9.4.2.1. Sabbatical programme*

Marie Doumic was in Vienna for a sabbatical stay until July 2018.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

- Jean Clairambault, with Jean-Frédéric Gerbeau, Huaxiong Huang, Benoît Perthame and Sivobal Sivaloganathan: organisers of the First Inria-Sorbonne-Université-Fields Institute international workshop on mathematics in medicine, Fields Institute, Toronto, January 31 to February 2, 2018. http://www.fields.utoronto.ca/activities/17-18/math-medicine
- Jean Clairambault, with Guillemette Chapuisat, Florence Hubert, Urszula Ledzewicz and Vitaly Volpert: organisers of the International 5-day workshop on Mathematical perspectives in the biology and therapeutics of cancer, CIRM Luminy, July 9-13, 2018. https://mathscancer.sciencesconf.org/
- Jean Clairambault, with Tomás Alarcón and Thomas Hillen: organisers of the BIRS-CMO 5-day workshop on Mathematical Challenges in the Analysis of Continuum Models for Cancer Growth, Evolution and Therapy (18w5115), Casa Matemática Oaxaca (Mexico), November 25-30, 2018. http://www.birs.ca/events/2018/5-day-workshops/18w5115
- Benoît Perthame, Luis Almeida, Diane Peurichard and Delphine Salort: organizers of the workshop 'Asymptotic approaches to spatial and dynamical organizations', july 2018, LJLL, Sorbonne Université, and organizers of the math-bio working group at LJLL, Sorbonne Université
- Diane Peurichard, Dirk Drasdo: organizers of the internal seminar 'Open MAMBA seminar' (https://team.inria.fr/mamba/open-mamba-seminar/)
- Marie Doumic, with Doron Levy (U. of Maryland, USA), M. Bergmann (Med. University of Vienna) and N. Mauser (Wolfgang Pauli Institute, Vienna) organised a workshop on "Mathematical Models in Cancer", on July 20th and 21st, 2018.
- Marie Doumic, together with J. Haskovec (KAUST Univ., Saoudi Arabia), M.-T. Wolfram (Univ. of Warwick, UK), K. Fellner (Univ. of Graz, Austria) and L. Neumann (U. of Innsbruck, Austria), organised a workshop on "Applied PDEs and kinetic equations: from physics to life sciences and beyond", in the honor of C. Schmeiser's 60th birthday.
- L. Almeida, B. Perthame, D. Peurichard and D. Salort have organized the international conference "Asymptotic approach to spatial and dynamical organizations", 4-6 july 2018 (60 participants)

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Member of the Conference Program Committees*

- Philippe Robert was in the PC of the conference "Stochastic Networks" held in Edimburgh 25-29 June 2018.
- Pierre-Alexandre Bliman was in the Editorial Board of the conference "European Control Conference", Limassol, 12-15 June 2018.
- Dirk Drasdo was member of the Program committee of the SBMC 2018 (Bremen, Germany), July 4-6, 2018.

*10.1.2.2. Reviewer*

- Pierre-Alexandre Bliman, reviewer for the conferences IEEE Conference on Decision Control, European Control Conference, Indian Control Conference, Joint 9th IFAC Symposium on Robust Control Design and 2nd IFAC Workshop on Linear Parameter Varying Systems

### *10.1.3. Journal*

*10.1.3.1. Member of the Editorial Boards*

- Philippe Robert is Associate Editor of the journal "Queueing Systems, Theory and Applications".
- Benoît Perthame is co-editor in chief of Acta Applicandae Mathematicae (Springer).
- Dirk Drasdo is associate editorial member of J. Theor. Biol. and Royal Society Open Science.
- Luis Almeida is Associate Editor of the Journal of Dynamics and Games AIMS.

*10.1.3.2. Reviewer - Reviewing Activities*

- Jean Clairambault, reviewer for different journals in 2018: PLoS Computational Biology, PLoS One, Integrative Biology, Applied Mathematical Modelling, Letters in Biomathematics, Journal of Theoretical Biology, Bulletin of Mathematical Biology, Mathematical Biosciences and Engineering
- Pierre-Alexandre Bliman, reviewer for the journals Automatica, Proceedings of the Royal Society of Edinburgh, Systems and Control Letters
- Diane Peurichard: reviewer for Journal of the Royal Society Interface, Processes, Computation
- Marie Doumic, reviewer for J. Math. Biol., SIAM App. Math., Ann. IHP, M2AN, Inverse Problems

### *10.1.4. Invited Talks*

- Jean Clairambault: Workshop "Computational Systems Biology for Cancer", Paris, January 2018; Spring school "Mathematical Modelling of Tumour Growth and Therapy", CRM, Bellaterra (UAB, Barcelona), April 2018; Workshop "Modelling Biological Phenomena from Nano to Macro Scales", The Fields Institute, Toronto, May 2018; 4th Annual workshop on mathematics in medicine: mathematical models in cancer, Wolfgang Pauli Institut, Vienna, July 2018; Seminar, DIMEAS, Politecnico di Torino, October 2018; Mathematical and Computational Biology Workshop, Tirana, October 2018; International Conference on Applied Mathematics, Fes (Morocco), October 2018 ; Seminar on mathematical modelling of tissue growth, Les Treilles (Tourtour, France), November 2018; Workshop on Mathematical Challenges in the Analysis of Continuum Models for Cancer Growth, Evolution and Therapy, CMO, Oaxaca, November 2018 (see above: organisation of scientific events); International Program in Mathematical Biology Sep.-Dec. 2018, Institut Mittag-Leffler (Stockholm), December 2018
- Diane Peurichard: Oct 16-18 PSU-P6 workshop "Mathematics of living matter", Penn State; Jul 23-Jul 27 11th ECMTB Conference, Lisbon, Portugal; Apr 30 May 04 Workshop "Collective dynamics and self-organization in biological sciences", Edinburgh, Scotland.
- Marie Doumic: Keynote speaker at the BIOMATH 2018 conference, Sofia, Bulgaria, June 25-29 2018; 3h Minicourse for the Doctoral School of Vienna, Weissensee, July 2-5; Workshop "Collective dynamics and self-organisation in biological sciences", Edinburgh, April 30-May 4
- Dirk Drasdo: Paris, France, Sanofi, 7/11, Seminar: Spatial-temporal MultiScale-multilevel modeling OF DRUG-INDUCED LIVER DAMAGE and ITS Regeneration : Steps towards a VIRTUAL Liver; Paris, France, Ecopa, 6/11, Symposium: how new experimental tools in life science challenge the 3R vision. (5/6.11): Spatial-temporal MultiScale-multilevel modeling of APAP damage and its consequence on ammonia detoxification : Steps towards a virtual Liver; Palaiseau, France, Ecole Polytechnique, 19/10, Guess lecture. Integrated and spatial-temporal multiscale modeling of liver guide in vivo experiments in healthy & chronic disease states: a blue print for systems medicine?; Rome, Italy, 10/10, Mathematics for Biomedicine, 8-11/10: Integrated and spatial-temporal multiscale modeling of liver guide in vivo experiments in healthy & chronic disease states: a blue print for systems medicine?; Pisa, Italy, 5/10, Multiscale modeling and simulations to bridge molecular and cellular scales, 1-5/10: Multi-scale and integrated mathematical modeling of liver function and regeneration: Using mathematical models to guide experiments; Heidelberg, Germany, 1/11, BIOMS-Symposium 2018, 1-2/11: Multi-scale and integrated mathematical modeling of liver function and regeneration: Using mathematical models to guide experiments; Inst. Pasteur, France 20/9, Inst. Seminar: From systems biology to systems medicine: an example: Paracetamol overdose-caused damage & its regeneration: Virtual experiments in organ micro-architectures; Paris, France, 14/4, EASL, 11-14/4: Parameterization of liver metabolism; Lyon, France : From systems biology to systems medicine: an example: Paracetamol overdose-caused damage & its regeneration: Virtual experiments in organ micro-architectures

### *10.1.5. Leadership within the Scientific Community*

Dirk Drasdo is member of the scientific leadership board of the German flagship project LiSyM (Liver Systems Medicine) financed by BMBF (Germany)

### *10.1.6. Scientific Expertise*

- Jean Clairambault and Dirk Drasdo: members of the ANR CES 45 (mathematics and digital sciences for health and biology) selection committee
- Jean Clairambault: member of the review committee for the German FZJ (ForschungsZentrum Jülich)-BMBF funding measure "Demonstrators for Individualised Medicine" within the Framework of the Research and Funding Concept "e:Med Ð Paving the Way for Systems Medicine", Frankfurt, September 2018
- Jean Clairambault: member of the review committee for the German DLR (Deutsches Zentrum für Luft und Raumfahrt)-BMBF initiative "Systems medicine research consortia", Berlin, November 2018
- Jean Clairambault: representative of Inria (until June 2018) to the expert group of the ITMO Cancer (Aviesan) and member of the steering committee of the HTE program (2016-2020)
- Diane Peurichard and Marie Doumic: Ambassadeur FSMP for Austria
- Diane Peurichard: member of the selection committee Sorbonne Université, maitre de conference position
- Diane Peurichard: member of the Inria scientific selection committee (CORDI-S PhD, post-docs, delegation)
- Benoît Perthame has been member of the "Chern Prize" committee awarded at ICM 2018 in Rio.

### *10.1.7. Research Administration*

- Pierre-Alexandre Bliman: Coordinator of the ECOS-Nord project C17M01 "News methods for controle of dengue and arboviroses epidemics".
- Marie Doumic: nominated in Dec. 2018 at the Scientific Council of INSMI, CNRS.
- Dirk Drasdo is member of the scientific leadership board of the German flagship project LiSyM (Liver Systems Medicine) financed by BMBF (Germany).
- Dirk Drasdo guides a research group bi-localized at Inria de Paris and IfADo, Dortmund, currently composed of 3 research engineers, 3 postdocs.

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master : Philippe Robert, "Large Stochastic Networks", 24h, M2, Sorbonne Université, France.

Master: Marie Doumic, "Inverse problems in Biological Dynamics", 24h, M2, Sorbonne Université, France

Jean Clairambault: teaching in Spring school "Mathematical Modelling of Tumour Growth and Therapy" (4 one-hour lectures), UAB, Barcelona, April 2018 (see above); Fall teaching (3 one-hour lectures), Politecnico di Torino, October 2018 (see above)

Licence : Pierre-Alexandre Bliman, "Calculus 3", 90h, L2 and "Analyse", 60h, L3, Fundação Getulio Vargas, Rio de Janeiro, Brazil

Master : Diane Peurichard, "Fondements des méthodes numériques", 20h, "Calcul matriciel numérique" 20h

Licence : Diane Peurichard, Licence: "Introduction to Python" 20h

Master : Dirk Drasdo, “Agent-based models of tissue organization”, UPMC M2 course, Paris **24 h / yr**

### *10.2.2. Supervision*

PhD : Wen Sun, “A study of interacting stochastic networks : large scale, long time behavior and fluctuations”, Sorbonne Université, June, 11, 2018, Philippe Robert

PhD : Veronica Quintuna Rodriguez, “New Network / IT Command: Virtualized Function Performance for a Programmable Infrastructure”, Sorbonne Université, October 3, 2018, Fabric Guillemin and Philippe Robert.

PhD in progress : Gaëtan Vignoud, “Plasticity of Stochastic Neural Networks”, September 1st, 2018, Philippe Robert, Laurent Venance

Jean Clairambault: supervision of the M2 (‘mathematics of modelling’) internship of Loïs Naudin: “Modélisation du métabolisme énergétique tumoral glycolytique vs. respiratoire oxydatif” and of two groups of four students of the L3 unit 3M101: “Excitabilité cellulaire : une première approche des systèmes dynamiques” and “Étude des équations de Lotka-Volterra”

Jean Clairambault: Supervision of PhD students Camille Pouchol (with Michèle Sabbah and Emmanuel Trélat, ED 386, Sorbonne Univerité, thesis defence June 2018: “Analysis, control and optimisation of PDEs, application to the biology and therapy of cancer” [1] and Ghassen Haddad (ED 386, Sorbonne Université, and Université Tunis-El Manar, co-supervisor: Slimane Ben Miled), thesis defence planned in December 2018: “Optimisation du traitement du cancer”)

Internship: Supervision of Valeria Caliaro, M1 student from University of Verona (3months internship) on interacting particle networks, Diane Peurichard

Internship: Supervision of S. Zhenyu, M1 student from LJLL, Sorbonne University (3months internship) on coarse graining of a fluid filled with obstacles, Diane Peurichard

Diane Peurichard, Luis Almeida, Benoit Perthame: supervision of project CEMRACS (6weeks)

PhD defended: Noémie Boissier PhD defence in June 2018, supervision by D. Drasdo and I. Vignon-Clementel

PhD in progress: Adrian Friebel, “Software of image processing and analysis of liver tissue at histological scales”, supervision by D. Drasdo and S. Hoehme

Internship: Ismael Gonzalez Valverde,PhD student from Zaragoza by Paul Van Liedekerke and Dirk Drasdo.

### *10.2.3. Juries*

- Ph.D thesis of C. Pouchol, defended on June 29th: participation in the committee of J. Clairambault and B. Perthame.
- Ph.D thesis of M. Strugarek, defended on September 7th: participation in the committee of M. Doumic and B. Perthame.
- Dirk Drasdo was in 2018 member of the ANR-grant selection committee for mathematical modeling in medicine and biology.

## 10.3. Popularization

### *10.3.1. Interventions*

- Marie Doumic: participation in a round table on December 13th, in Forum Emploi Maths, La Villette, Paris; Participation in a round table on ”Science Meets Medicine”, May 29th, in Vienna; presentation to a 12-year old class of pupils of mathematics for biology.
- Several Mamba members have taken part to the activities of the year of mathematical biology declared by European Mathematical Society.

- Dirk Drasdo on liver research and EASL: https://www.inria.fr/centre/paris/actualites/dirk-drasdo-et-ses-recherches-sur-le-foie

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] C. POUCHOL.*Analysis, control and optimisation of PDEs, application to the biology and therapy of cancer*, Sorbonne Université, June 2018, https://hal.archives-ouvertes.fr/tel-01889253

[2] V. K. QUINTUNA RODRIGUEZ.*New Network / IT Command: Virtualized Function Performance for a Programmable Infrastructure*, Sorbonne Université, October 2018, https://hal.inria.fr/tel-01884431

[3] M. STRUGAREK.*Mathematical modeling of population dynamics, applications to vector control of Aedes spp. (Diptera:Culicidae)*, Sorbonne Université , UPMC, September 2018, https://hal.archives-ouvertes.fr/tel-01879201

[4] W. SUN.*A study of interacting stochastic networks :large scale, long time behavior and fluctuations*, Université Pierre et Marie Curie, June 2018, https://hal.inria.fr/tel-01813434

[5] C. TAING.*Concentration dynamics in non local partial differential equations from biology*, Sorbonne Université, UPMC University of Paris 6, Laboratoire Jacques-Louis Lions, June 2018, https://tel.archives-ouvertes.fr/tel-01922892

### Articles in International Peer-Reviewed Journal

[6] R. AGHAJANI, P. ROBERT, W. SUN.*A Large Scale Analysis of Unreliable Stochastic Networks*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", April 2018, vol. 28, n$^{o}$ 2, 36, https://arxiv.org/abs/1608.08743 [*DOI :* 10.1214/17-AAP1318], https://hal.archives-ouvertes.fr/hal-01359208

[7] V. CALVEZ, B. PERTHAME, S. YASUDA.*Traveling Wave and Aggregation in a Flux-Limited Keller-Segel Model*, in "Kinetic and Related Models ", 2018, vol. 11, n$^{o}$ 4, p. 891–909, https://arxiv.org/abs/1709.07296 [*DOI :* 10.3934/KRM.2018035], https://hal.sorbonne-universite.fr/hal-01591490

[8] J. CALVO, M. DOUMIC, B. PERTHAME.*Long-time asymptotics for polymerization models*, in "Communications in Mathematical Physics", October 2018, vol. 363, n$^{o}$ 1, p. 111-137, https://arxiv.org/abs/1707.09777 [*DOI :* 10.1007/S00220-018-3218-5], https://hal.archives-ouvertes.fr/hal-01570292

[9] T. CHAGAS, P.-A. BLIMAN, K. KIENITZ.*Stabilization of periodic orbits of discrete-time dynamical systems using the Prediction-Based Control: New control law and practical aspects*, in "Journal of The Franklin Institute", August 2018, vol. 355, n$^{o}$ 12, p. 4771-4793 [*DOI :* 10.1016/J.JFRANKLIN.2018.04.040], https://hal.inria.fr/hal-01941693

[10] T. P. CHAGAS, P.-A. BLIMAN, K. H. KIENITZ.*Approximate Prediction-Based Control Method for Nonlinear Oscillatory Systems with Applications to Chaotic Systems*, in "Journal of Control Science and Engineering", March 2018, vol. 2018, p. 1-29 [*DOI :* 10.1155/2018/3298286], https://hal.inria.fr/hal-01722175

[11] J. CLAIRAMBAULT, B. PERTHAME, A. QUILLAS MARAN.*Analysis of a system describing proliferative-quiescent cell dynamics*, in "Chinese Annals of Mathematics - Series B", 2018, p. 1-13, https://hal.sorbonne-universite.fr/hal-01674142

[12] W. DJEMA, C. BONNET, F. MAZENC, J. CLAIRAMBAULT, E. FRIDMAN, P. HIRSCH, F. DELHOMMEAU.*Control in dormancy or eradication of cancer stem cells: Mathematical modeling and stability issues*, in "Journal of Theoretical Biology", July 2018, vol. 449, p. 103 - 123 [*DOI :* 10.1016/J.JTBI.2018.03.038], https://hal.inria.fr/hal-01852154

[13] M. DOUMIC, M. ESCOBEDO, M. TOURNUS.*Estimating the division rate and kernel in the fragmentation equation*, in "Annales de l'Institut Henri Poincaré (C) Non Linear Analysis", 2018, https://arxiv.org/abs/1804.08945 [*DOI :* 10.1016/J.ANIHPC.2018.03.004], https://hal.archives-ouvertes.fr/hal-01501811

[14] A. GHALLAB, U. HOFMANN, S. SEZGIN, N. VARTAK, R. HASSAN, A. ZAZA, P. GODOY, K. M. SCHNEIDER, G. GUENTHER, Y. A. AHMED, A. ABBAS, V. KEITEL, L. KUEPFER, S. DOOLEY, F. LAMMERT, C. TRAUTWEIN, M. SPITELLER, D. DRASDO, A. HOFMANN, P. L. JANSEN, J. HENGSTLER, R. REIF.*Bile Microinfarcts in Cholestasis Are Initiated by Rupture of the Apical Hepatocyte Membrane and Cause Shunting of Bile to Sinusoidal Blood*, in "Hepatology", August 2018, https://hal.inria.fr/hal-01968843

[15] S. HOEHME, F. BERTAUX, W. WEENS, B. GRASL-KRAUPP, J. HENGSTLER, D. DRASDO.*Model Prediction and Validation of an Order Mechanism Controlling the Spatiotemporal Phenotype of Early Hepatocellular Carcinoma*, in "Bulletin of Mathematical Biology", May 2018, vol. 80, n$^{o}$ 5, p. 1134-1171, https://hal.inria.fr/hal-01968844

[16] L. NEVES DE ALMEIDA, F. BUBBA, B. PERTHAME, C. POUCHOL.*Energy and implicit discretization of the Fokker-Planck and Keller-Segel type equations*, in "Networks and Heterogeneous Media", March 2019, vol. 14, n$^{o}$ 1, https://arxiv.org/abs/1803.10629 [*DOI :* 10.3934/NHM.2019002], https://hal.archives-ouvertes.fr/hal-01745769

[17] S. NORDMANN, B. PERTHAME, C. TAING.*Dynamics of concentration in a population model structured by age and a phenotypical trait*, in "Acta Applicandae Mathematicae", June 2018, vol. 155, n$^{o}$ 1 [*DOI :* 10.1007/S10440-017-0151-0], https://hal.archives-ouvertes.fr/hal-01493068

[18] G. PEETERS, C. DEBBAUT, A. FRIEBEL, P. CORNILLIE, W. DE VOS, K. FAVERE, I. VANDER ELST, T. VANDECASTEELE, T. JOHANN, L. VAN HOOREBEKE, D. MONBALIU, D. DRASDO, S. HOEHME, W. LALEMAN, P. SEGERS.*Quantitative analysis of hepatic macro- and microvascular alterations during cirrhogenesis in the rat*, in "Journal of Anatomy", 2018 [*DOI :* 10.1111/JOA.12760], https://hal.inria.fr/hal-01700102

[19] B. PERTHAME, E. RIBES, D. SALORT.*Career plans and wage structures: a mean field game approach*, in "Mathematics in Engineering", May 2018, vol. 1, n$^{o}$ 1, p. 47-63 [*DOI :* 10.3934/MINE.2018.1.47], https://hal.sorbonne-universite.fr/hal-01674630

[20] B. PERTHAME, W. SUN, M. TANG.*The fractional diffusion limit of a kinetic model with biochemical pathway*, in "Zeitschrift für Angewandte Mathematik und Physik", May 2018, vol. 69:67, https://arxiv.org/abs/1709.03308 [*DOI :* 10.1007/S00033-018-0964-3], https://hal.sorbonne-universite.fr/hal-01584754

[21] B. PERTHAME, S. YASUDA.*Stiff-response-induced instability for chemotactic bacteria and flux-limited Keller-Segel equation*, in "Nonlinearity", 2018, vol. 31, n$^{o}$ 9 [*DOI :* 10.1088/1361-6544/AAC760], https://hal.sorbonne-universite.fr/hal-01494963

[22] C. POUCHOL, J. CLAIRAMBAULT, A. LORZ, E. TRÉLAT.*Asymptotic analysis and optimal control of an integro-differential system modelling healthy and cancer cells exposed to chemotherapy*, in "Journal de Mathématiques Pures et Appliquées", 2018, vol. 116, p. 268–308, https://arxiv.org/abs/1612.04698 [*DOI :* 10.1016/J.MATPUR.2017.10.007], https://hal.archives-ouvertes.fr/hal-01673589

[23] C. POUCHOL, E. TRÉLAT.*Global stability with selection in integro-differential Lotka-Volterra systems modelling trait-structured populations*, in "Journal of Biological Dynamics", 2018, vol. 12, n$^{o}$ 1, p. 872–893, https://arxiv.org/abs/1702.06187 , https://hal.archives-ouvertes.fr/hal-01470722

[24] P. ROBERT, A. VEBER.*A Scaling Analysis of a Star Network with Logarithmic Weights*, in "Stochastic Processes and their Applications", June 2018, https://arxiv.org/abs/1609.04180 , https://hal.inria.fr/hal-01377703

[25] W. SUN.*A Functional Central Limit Theorem for the Becker-Döring model*, in "Journal of Statistical Physics", 2018, vol. 171, n$^{o}$ 1, p. 145–165, https://arxiv.org/abs/1710.04059 - 18 pages [*DOI :* 10.04059], https://hal.inria.fr/hal-01616039

### International Conferences with Proceedings

[26] M. S. ARONNA, P.-A. BLIMAN.*Interval observer for uncertain time-varying SIR-SI epidemiological model of vector-borne disease*, in "ECC 2018 - European Control Conference", Limassol, Cyprus, June 2018, p. 1-6, https://arxiv.org/abs/1703.07083 , https://hal.inria.fr/hal-01493078

[27] 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

[28] W. DJEMA, C. BONNET, F. MAZENC, J. CLAIRAMBAULT.*Introducing Cell-Plasticity Mechanisms into a Class of Cell Population Dynamical Systems*, in "IEEE American Control Conference (ACC 2018)", Milwaukee, United States, June 2018 [*DOI :* 10.23919/ACC.2018.8430758], https://hal.inria.fr/hal-01848890

[29] W. DJEMA, F. MAZENC, C. BONNET, J. CLAIRAMBAULT, E. FRIDMAN.*Stability Analysis of a Nonlinear System with Infinite Distributed Delays Describing Cell Dynamics*, in "IEEE American Control Conference (ACC 2018)", Milwaukee, United States, June 2018 [*DOI :* 10.23919/ACC.2018.8430869], https://hal.inria.fr/hal-01849010

### Scientific Books (or Scientific Book chapters)

[30] D. DRASDO, A. BUTTENSCHÖN, P. VAN LIEDEKERKE.*Agent-Based Lattice Models of Multicellular Systems*, in "Numerical Methods and Advanced Simulation in Biomechanics and Biological Processes", Elsevier, 2018, p. 223-238 [*DOI :* 10.1016/B978-0-12-811718-7.00012-5], https://hal.inria.fr/hal-01968847

[31] L. NEVES DE ALMEIDA, R. H. CHISHOLM, J. CLAIRAMBAULT, T. LORENZI, A. LORZ, C. POUCHOL, E. TRÉLAT.*Why Is Evolution Important in Cancer and What Mathematics Should Be Used to Treat Cancer? Focus on Drug Resistance*, in "Trends in Biomathematics: Modeling, Optimization and Computational Problems: Selected works from the BIOMAT Consortium Lectures, Moscow 2017", Springer International Publishing, August 2018, p. 107-120, https://hal.inria.fr/hal-01945593

[32] P. VAN LIEDEKERKE, A. BUTTENSCHÖN, D. DRASDO.*Off-Lattice Agent-Based Models for Cell and Tumor Growth*, in "Numerical Methods and Advanced Simulation in Biomechanics and Biological Processes", Elsevier, 2018, p. 245-267 [*DOI :* 10.1016/B978-0-12-811718-7.00014-9], https://hal.inria.fr/hal-01968846

## Books or Proceedings Editing

[33] M. DOUMIC, B. VAN BRUNT (editors). *Explicit Solution and Fine Asymptotics for a Critical Growth-Fragmentation Equation*, EDP Sciences, October 2018, vol. 62, p. 30-42, https://arxiv.org/abs/1704.06087 [*DOI :* 10.1051/PROC/201862030], https://hal.archives-ouvertes.fr/hal-01510960

## Other Publications

[34] P.-A. BLIMAN.*Feedback Control Principles for Biological Control of Dengue Vectors*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01944958

[35] P.-A. BLIMAN, D. CARDONA-SALGADO, Y. DUMONT, O. VASILIEVA.*Implementation of Control Strategies for Sterile Insect Techniques*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01943683

[36] F. BUBBA, B. PERTHAME, C. POUCHOL, M. SCHMIDTCHEN.*Hele-Shaw limit for a system of two reaction-(cross-)diffusion equations for living tissues*, January 2019, https://arxiv.org/abs/1901.01692 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01970313

[37] F. BUBBA, C. POUCHOL, N. FERRAND, G. VIDAL, L. NEVES DE ALMEIDA, B. PERTHAME, M. SABBAH.*A chemotaxis-based explanation of spheroid formation in 3D cultures of breast cancer cells*, October 2018, https://arxiv.org/abs/1810.13162 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01895999

[38] D. BÉAL, M. TOURNUS, R. MARCHANTE, T. PURTON, D. SMITH, M. F. TUITE, M. DOUMIC, W.-F. XUE.*The Division of Amyloid Fibrils*, December 2018, working paper or preprint [*DOI :* 10.1101/506386], https://hal.archives-ouvertes.fr/hal-01966243

[39] M. DOUMIC, K. FELLNER, M. MEZACHE, H. REZAEI.*A bi-monomeric, nonlinear Becker-Döring-type system to capture oscillatory aggregation kinetics in prion dynamics*, August 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01863748

[40] T. DĘBIEC, M. DOUMIC, P. GWIAZDA, E. WIEDEMANN.*Relative Entropy Method for Measure Solutions of the Growth-Fragmentation Equation*, April 2018, https://arxiv.org/abs/1804.03538 - working paper or preprint, https://hal.inria.fr/hal-01762974

[41] D. FANG, S. JIN, P. MARKOWICH, B. PERTHAME.*Implicit and Semi-implicit Numerical Schemes for the Gradient Flow of the Formation of Biological Transport Networks*, January 2019, working paper or preprint, https://hal.sorbonne-universite.fr/hal-01984371

[42] C. HENDERSON, B. PERTHAME, P. E. SOUGANIDIS.*Super-linear propagation for a general, local cane toads model*, May 2018, https://arxiv.org/abs/1705.04029 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01520145

[43] P. V. LIEDEKERKE, J. NEITSCH, T. JOHANN, K. ALESSANDRI, P. NASSOY, D. DRASDO.*Quantitative agent-based modeling reveals mechanical stress response of growing tumor spheroids is predictable over various growth conditions and cell lines*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01956017

[44] P. V. LIEDEKERKE, J. NEITSCH, T. JOHANN, E. WARMT, I. GONZÀLEZ-VALVERDE, S. HOEHME, S. GROSSER, J. KAES, D. DRASDO.*A quantitative high resolution computational mechanics cell model for growing and regenerating tissues*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01956023

[45] D. MARTIROSYAN, V. NERSESYAN.*Multiplicative ergodic theorem for a non-irreducible random dynamical system*, January 2018, https://arxiv.org/abs/1801.09440 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01695046

[46] D. MARTIROSYAN, P. ROBERT.*The Equilibrium States of Large Networks of Erlang Queues*, November 2018, https://arxiv.org/abs/1811.04763 - working paper or preprint, https://hal.inria.fr/hal-01920171

[47] L. NAUDIN.*Modélisation du métabolisme énergétique tumoral glycolytique vs. respiratoire oxydatif*, Laboratoire Jacques-Louis Lions, UPMC, Paris, September 2018, https://hal.inria.fr/hal-01980569

[48] L. NEVES DE ALMEIDA, M. DUPREZ, Y. PRIVAT, N. VAUCHELET.*Control strategies on mosquitos population for the fight against arboviruses*, January 2019, https://arxiv.org/abs/1901.05688 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01984426

[49] L. NEVES DE ALMEIDA, Y. PRIVAT, M. STRUGAREK, N. VAUCHELET.*Optimal releases for population replacement strategies, application to Wolbachia*, June 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01807624

[50] T. N. NGUYEN, J. CLAIRAMBAULT, T. JAFFREDO, B. PERTHAME, D. SALORT.*Adaptive dynamics of hematopoietic stem cells and their supporting stroma: A model and mathematical analysis*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01963820

[51] A. OLIVIER, C. POUCHOL.*Combination of direct methods and homotopy in numerical optimal control: application to the optimization of chemotherapy in cancer*, January 2018, https://arxiv.org/abs/1707.08038 - working paper or preprint, https://hal-auf.archives-ouvertes.fr/hal-01568779

[52] B. PERTHAME, D. SALORT.*Derivation of an integrate&fire equation for neural networks from a voltage-conductance kinetic model*, December 2018, working paper or preprint, https://hal.sorbonne-universite.fr/hal-01881950

[53] B. PERTHAME, N. VAUCHELET, Z. WANG.*The Flux Limited Keller-Segel System; Properties and Derivation from Kinetic Equations*, January 2018, https://arxiv.org/abs/1801.07062 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01689571

[54] B. PICCOLI, N. POURADIER DUTEIL, E. TRÉLAT.*Sparse control of Hegselmann-Krause models: Black hole and declustering*, February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01699261

[55] C. POUCHOL.*On the stability of the state 1 in the non-local Fisher-KPP equation in bounded domains*, January 2018, https://arxiv.org/abs/1801.05653 - working paper or preprint, https://hal.sorbonne-universite.fr/hal-01686461

[56] C. POUCHOL, E. TRÉLAT, E. ZUAZUA.*Phase portrait control for 1D monostable and bistable reaction-diffusion equations*, May 2018, https://arxiv.org/abs/1805.10786 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01800382

[57] M. STRUGAREK, H. BOSSIN, Y. DUMONT.*On the use of the sterile insect technique or the incompatible insect technique to reduce or eliminate mosquito populations*, May 2018, https://arxiv.org/abs/1805.10150 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01799954

[58] M. STRUGAREK, L. DUFOUR, N. VAUCHELET, L. NEVES DE ALMEIDA, B. PERTHAME, D. A. M. VILLELA.*Oscillatory regimes in a mosquito population model with larval feedback on egg hatching*, January 2018, https://arxiv.org/abs/1801.03701 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01674280

[59] M. STRUGAREK, H. JI.*Sharp seasonal threshold property for cooperative population dynamics with concave nonlinearities*, April 2018, https://arxiv.org/abs/1804.07641 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01772628

[60] W. SUN, P. ROBERT.*Analysis of Large Urn Models with Local Mean-Field Interactions*, February 2018, https://arxiv.org/abs/1802.05064 - working paper or preprint, https://hal.inria.fr/hal-01710964

## References in notes

[61] L. ALPHEY.*Genetic control of mosquitoes*, in "Annual review of entomology", 2014, vol. 59

[62] A. ARMIENTO.*Inverse problems and data assimilation methods applied to protein polymerisation*, Université Paris 7 - Diderot, January 2017, https://hal.inria.fr/tel-01447286

[63] A. ARMIENTO, M. DOUMIC, P. MOIREAU, H. REZAEI.*Estimation from Moments Measurements for Amyloid Depolymerisation*, in "Journal of Theoretical Biology", March 2016 [*DOI :* 10.1016/J.JTBI.2016.02.037], https://hal.archives-ouvertes.fr/hal-01248255

[64] A. ARMIENTO, P. MOIREAU, D. MARTIN, N. LEPEJOVA, M. DOUMIC, H. REZAEI.*The mechanism of monomer transfer between two structurally distinct PrP oligomers*, in "PLoS ONE", July 2017, vol. 12, n$^{o}$ 7 [*DOI :* 10.1371/JOURNAL.PONE.0180538], https://hal.archives-ouvertes.fr/hal-01574346

[65] J. L. AVILA ALONSO, C. BONNET, J. CLAIRAMBAULT, H. OZBAY, S.-I. NICULESCU, F. MERHI, A. BALLESTA, R. TANG, J.-P. MARIE.*Analysis of a New Model of Cell Population Dynamics in Acute Myeloid Leukemia*, in "Delay Systems : From Theory to Numerics and Applications", T. VYHLÍDAL, J.-F. LAFAY, R. SIPAHI (editors), Advances in Delays and Dynamics, Springer, January 2014, vol. 1, p. 315-328 [*DOI :* 10.1007/978-3-319-01695-5_23], https://hal.inria.fr/hal-00932779

[66] J. L. AVILA ALONSO, C. BONNET, E. FRIDMAN, F. MAZENC, J. CLAIRAMBAULT.*Stability analysis of PDE's modelling cell dynamics in Acute Myeloid Leukemia*, in "53rd IEEE Conference on Decision and Control", Los Angeles, United States, December 2014, https://hal.inria.fr/hal-01110304

[67] J. L. AVILA ALONSO, C. BONNET, H. OZBAY, J. CLAIRAMBAULT, S.-I. NICULESCU.*A coupled model for healthy and cancer cells dynamics in Acute Myeloid Leukemia*, in "The 19th World Congress of the International Federation of Automatic Control", Cape Town, Souh Africa, August 2014, https://hal.inria.fr/hal-00940245

[68] J. L. AVILA ALONSO, C. BONNET, H. OZBAY, J. CLAIRAMBAULT, S.-I. NICULESCU.*A discrete-maturity Interconnected Model of Healthy and Cancer Cell Dynamics in Acute Myeloid Leukemia*, in "Mathematical Theory of Networks and Systems", Groningen, Netherlands, July 2014, https://hal.inria.fr/hal-00940305

[69] H. T. BANKS, M. DOUMIC, C. KRUSE.*A numerical scheme for the early steps of nucleation-aggregation Models*, in "Journal of Mathematical Biology", January 2017, vol. 74, n$^{o}$ 1-2, p. 259-287 [*DOI :* 10.1007/s00285-016-1026-0], https://hal.inria.fr/hal-00954437

[70] H. T. BANKS, M. DOUMIC, C. KRUSE, S. PRIGENT, H. REZAEI.*Information Content in Data Sets for a Nucleated-Polymerization Model*, in "Journal of Biological Dynamics", June 2015, vol. 9, n$^{o}$ 1, 26 [*DOI :* 10.1080/17513758.2015.1050465], https://hal.inria.fr/hal-01123847

[71] J. BARRÉ, J. A. CARRILLO, P. DEGOND, E. ZATORSKA, D. PEURICHARD. *Particle interactions mediated by dynamical networks: assessment of macroscopic descriptions*, in " Journal of Nonlinear Science", 2017

[72] J. BARRÉ, P. DEGOND, A. KHUONG, E. ZATORSKA, D. PEURICHARD. *A two-species macroscopic model for cell segregation and border sharpening by Eph receptor ephrin-mediated repulsion*, in "to be submitted", 2019

[73] F. BEKKAL BRIKCI, J. CLAIRAMBAULT, B. PERTHAME.*Analysis of a molecular structured population model with possible polynomial growth for the cell division cycle*, in "Math. Comput. Modelling", 2008, vol. 47, n$^{o}$ 7-8, p. 699–713

[74] F. BERTAUX, S. STOMA, D. DRASDO, G. BATT.*Modeling Dynamics of Cell-to-Cell Variability in TRAIL-Induced Apoptosis Explains Fractional Killing and Predicts Reversible Resistance*, in "PLoS Computational Biology", 2014, vol. 10, n$^{o}$ 10, 14 [*DOI :* 10.1371/JOURNAL.PCBI.1003893.S016], https://hal.inria.fr/hal-00942885

[75] J. BERTOIN, A. R. WATSON.*Probabilistic aspects of critical growth-fragmentation equations*, in "Advances in Applied Probability", 9 2015

[76] P.-A. BLIMAN, M. S. ARONNA, F. C. COELHO, M. A. H. B. DA SILVA.*Ensuring successful introduction of Wolbachia in natural populations of Aedes aegypti by means of feedback control*, in "Journal of Mathematical Biology", August 2017 [*DOI :* 10.1007/s00285-017-1174-x], https://hal.inria.fr/hal-01579477

[77] C. BONNET, J. L. AVILA ALONSO, H. OZBAY, J. CLAIRAMBAULT, S.-I. NICULESCU, P. HIRSCH.*A Discrete-Maturity Interconnected Model of Healthy and Cancer Cell Dynamics in Acute Myeloid Leukemia*, in "The 10th AIMS Conference on Dynamical Systems,Differential Equations and Applications", Madrid, Spain, July 2014, https://hal.inria.fr/hal-01110309

[78] T. BOURGERON, M. DOUMIC, M. ESCOBEDO.*Estimating the division rate of the growth-fragmentation equation with a self-similar kernel*, in "Inverse Problems", Jan 2014, vol. 30, n$^{o}$ 2, 025007, http://dx.doi.org/10.1088/0266-5611/30/2/025007

[79] T. BOURGERON, Z. XU, M. DOUMIC, M. T. TEIXEIRA.*The asymmetry of telomere replication contributes to replicative senescence heterogeneity*, in "Scientific Reports", October 2015, vol. 5, 15326 [*DOI :* 10.1038/SREP15326], http://hal.upmc.fr/hal-01272075

[80] K. BOURTZIS.*Wolbachia-Based Technologies for Insect Pest Population Control*, in "Advances in Experimental Medicine and Biology", Springer, New York, NY, 02 2008, vol. 627

[81] A. BROWN.*Insecticide resistance in mosquitoes: a pragmatic review.*, in "Journal of the American Mosquito Control Association", 1986, vol. 2, n$^{o}$ 2, p. 123–140

[82] P. O. BUCUR, M. BEKHEIT, C. AUDEBERT, A. OTHMAN, S. HAMMAD, M. SEBAGH, M. A. ALLARD, B. DECANTE, A. FRIEBEL, D. DRASDO, E. MIQUELESTORENA-STANDLEY, J. G. HENGSTLER, I. VIGNON-CLEMENTEL, E. VIBERT.*Modulating Portal Hemodynamics With Vascular Ring Allows Efficient Regeneration After Partial Hepatectomy in a Porcine Model.*, in "Annals of Surgery", February 2017 [*DOI :* 10.1097/SLA.0000000000002146], https://hal.archives-ouvertes.fr/hal-01494844

[83] M. J. CACERES, B. PERTHAME.*Beyond blow-up in excitatory integrate and fire neuronal networks: refractory period and spontaneous activity*, in "Journal of Theoretical Biology", 2014, vol. 350, p. 81-89 [*DOI :* 10.1016/J.JTBI.2014.02.005], http://hal.upmc.fr/hal-00874746

[84] V. CALVEZ, M. DOUMIC, P. GABRIEL.*Self-similarity in a general aggregation–fragmentation problem. Application to fitness analysis*, in "Journal de Mathématiques Pures et Appliquées", 2012, vol. 98, n$^{o}$ 1, p. 1 - 27 [*DOI :* 10.1016/J.MATPUR.2012.01.004], http://www.sciencedirect.com/science/article/pii/S002178241200013X

[85] V. CALVEZ, N. LENUZZA, M. DOUMIC, J.-P. DESLYS, F. MOUTHON, B. PERTHAME.*Prion dynamic with size dependency - strain phenomena*, in "J. of Biol. Dyn.", 2010, vol. 4, n$^{o}$ 1, p. 28–42

[86] J. A. CARRILLO, F. JAMES, F. LAGOUTIÈRE, N. VAUCHELET.*The Filippov characteristic flow for the aggregation equation with mildly singular potentials*, in "Journal of Differential Equations", 2016, vol. 260, n$^{o}$ 1, p. 304-338, 33 pages, https://hal.archives-ouvertes.fr/hal-01061991

[87] J. CHEVALLIER, M. J. CACERES, M. DOUMIC, P. REYNAUD-BOURET.*Microscopic approach of a time elapsed neural model*, in "Mathematical Models and Methods in Applied Sciences", December 2015, 2669 [*DOI :* 10.1142/S021820251550058X], http://hal.upmc.fr/hal-01159215

[88] R. H. CHISHOLM, T. LORENZI, J. CLAIRAMBAULT.*Cell population heterogeneity and evolution towards drug resistance in cancer: Biological and mathematical assessment, theoretical treatment optimisation*, in "BBA - General Subjects", June 2016, vol. 1860, p. 2627 - 2645 [*DOI :* 10.1016/J.BBAGEN.2016.06.009], https://hal.inria.fr/hal-01321535

[89] R. H. CHISHOLM, T. LORENZI, A. LORZ, A. K. LARSEN, L. NEVES DE ALMEIDA, A. ESCARGUEIL, J. CLAIRAMBAULT.*Emergence of Drug Tolerance in Cancer Cell Populations: An Evolutionary Outcome of Selection, Nongenetic Instability, and Stress-Induced Adaptation*, in "Cancer Research", March 2015, vol. 75, n$^{o}$ 6, p. 930-939 [*DOI :* 10.1158/0008-5472.CAN-14-2103], https://hal.inria.fr/hal-01237893

[90] R. H. CHISHOLM, T. LORENZI, A. LORZ, A. K. LARSEN, L. NEVES DE ALMEIDA, A. ESCARGUEIL, J. CLAIRAMBAULT.*Emergence of Drug Tolerance in Cancer Cell Populations: An Evolutionary Outcome of Selection, Nongenetic Instability, and Stress-Induced Adaptation*, in "Cancer Research", Jan 2015, vol. 75, n$^{o}$ 6, p. 930–939 [*DOI :* 10.1158/0008-5472.CAN-14-2103], http://dx.doi.org/10.1158/0008-5472.CAN-14-2103

[91] W. DJEMA, F. MAZENC, C. BONNET, J. CLAIRAMBAULT, P. HIRSCH, F. DELHOMMEAU.*Stability of a Delay System Coupled to a Differential-Difference System Describing the Coexistence of Ordinary and Mutated Hematopoietic Stem Cells*, in "Conference on Decision and Control ", Las Vegas, United States, December 2016, https://hal.inria.fr/hal-01389870

[92] M. DOUMIC, M. ESCOBEDO.*Time Asymptotics for a Critical Case in Fragmentation and Growth-Fragmentation Equations*, in "Kinetic and Related Models ", June 2016, vol. 9, n$^{o}$ 2, 47 [*DOI :* 10.3934/KRM.2016.9.251], https://hal.inria.fr/hal-01080361

[93] M. DOUMIC, S. EUGENE, P. ROBERT.*Asymptotics of Stochastic Protein Assembly Models*, in "SIAM Journal on Applied Mathematics", November 2016, vol. 76, n$^{o}$ 6, 20 [*DOI :* 10.1137/16M1066920], https://hal.inria.fr/hal-01301266

[94] M. DOUMIC, P. GABRIEL.*Eigenelements of a General Aggregation-Fragmentation Model*, in "Mathematical Models and Methods in Applied Sciences", 2009, vol. 20, n$^{o}$ 05, 757, http://arxiv.org/abs/0907.5467

[95] M. DOUMIC, M. HOFFMANN, N. KRELL, L. ROBERT.*Statistical estimation of a growth-fragmentation model observed on a genealogical tree*, October 2012, 46 pages, 4 figures, https://hal.archives-ouvertes.fr/hal-00763601

[96] M. DOUMIC, B. PERTHAME, J. ZUBELLI.*Numerical Solution of an Inverse Problem in Size-Structured Population Dynamics*, in "Inverse Problems", 2009, vol. 25, n$^{o}$ 4, 045008

[97] D. DRASDO, S. HOEHME, J. G. HENGSTLER.*How predictive quantitative modeling of tissue organization can inform liver disease pathogenesis*, in "Journal of Hepatology", October 2014, vol. 61, n$^{o}$ 4, p. 951–956 [*DOI :* 10.1016/J.JHEP.2014.06.013], https://hal.inria.fr/hal-01110644

[98] V. A. DYCK, J. HENDRICHS, A. S. ROBINSON.*The Sterile Insect Technique, Principles and Practice in Area-Wide Integrated Pest Management*, Springer, Dordrecht, 2006, 787

[99] S. EUGENE.*Stochastic modelling in molecular biology: a probabilistic analysis of protein polymerisation and telomere shortening*, UPMC LJLL, September 2016, https://hal.inria.fr/tel-01377561

[100] S. EUGENE, W.-F. XUE, P. ROBERT, M. DOUMIC.*Insights into the variability of nucleated amyloid polymerization by a minimalistic model of stochastic protein assembly*, in "Journal of Chemical Physics", May 2016, vol. 144, n$^{o}$ 17, 12 [*DOI :* 10.1063/1.4947472], https://hal.inria.fr/hal-01205549

[101] A. GHALLAB, G. CELLIÈRE, S. HENKEL, D. DRIESCH, S. HOEHME, U. HOFMANN, S. ZELLMER, P. GODOY, A. SACHINIDIS, M. BLASZKEWICZ, R. REIF, R. MARCHAN, L. KUEPFER, D. HÄUSSINGER, D. DRASDO, G. GEBHARDT, J. G. HENGSTLER.*Model-guided identification of a therapeutic strategy to reduce hyperammonemia in liver diseases*, in "Journal of Hepatology", November 2015, vol. 64, n$^{o}$ 4, p. 860–871 [*DOI :* 10.1016/J.JHEP.2015.11.018], https://hal.archives-ouvertes.fr/hal-01257127

[102] L. GOSSE, N. VAUCHELET.*Hydrodynamic singular regimes in 1+1 kinetic models and spectral numerical methods*, in "Journal of Mathematical Analysis and Applications", 2016 [*DOI :* 10.1016/J.JMAA.2016.07.059], https://hal.archives-ouvertes.fr/hal-01354980

[103] J. HEMINGWAY, H. RANSON.*Insecticide resistance in insect vectors of human disease*, in "Annual review of entomology", 2000, vol. 45, n$^{o}$ 1, p. 371–391

[104] M. HERTIG, S. B. WOLBACH.*Studies on rickettsia-like micro-organisms in insects*, in "The Journal of medical research", 1924, vol. 44, n$^{o}$ 3, 329

[105] V. H. HOANG.*Adaptive estimation for inverse problems with applications to cell divisions*, Université de Lille 1 – Sciences et Technologies, November 2016, https://tel.archives-ouvertes.fr/tel-01417780

[106] M. HOFFMANN, A. OLIVIER.*Nonparametric estimation of the division rate of an age dependent branching process*, in "Stochastic Processes and their Applications", December 2015 [*DOI :* 10.1016/J.SPA.2015.11.009], https://hal.archives-ouvertes.fr/hal-01254203

[107] N. JAGIELLA, B. MÜLLER, M. MÜLLER, I. E. VIGNON-CLEMENTEL, D. DRASDO.*Inferring Growth Control Mechanisms in Growing Multi-cellular Spheroids of NSCLC Cells from Spatial-Temporal Image Data*, in "PLoS Computational Biology", 2016, vol. 12, n$^{o}$ 2, e1004412 [*DOI :* 10.1371/JOURNAL.PCBI.1004412], http://hal.upmc.fr/hal-01244593

[108] F. JAMES, N. VAUCHELET.*Numerical methods for one-dimensional aggregation equations*, in "SIAM Journal on Numerical Analysis", 2015, vol. 53, n$^{o}$ 2, p. 895-916 [*DOI :* 10.1137/140959997], https://hal.archives-ouvertes.fr/hal-00955971

[109] F. JAMES, N. VAUCHELET.*Equivalence between duality and gradient flow solutions for one-dimensional aggregation equations*, in "Discrete and Continuous Dynamical Systems - Series A", 2016, vol. 36, n$^{o}$ 3, p. 1355-1382, https://hal.archives-ouvertes.fr/hal-00803709

[110] G. JANKOWIAK, D. PEURICHARD, A. REVERSAT, M. SIXT, C. SCHMEISER.*Modelling adhesion-independent cell migration*, in "to be submitted", 2019

[111] M.-J. KANG, B. PERTHAME, D. SALORT.*Dynamics of time elapsed inhomogeneous neuron network model*, in "Comptes Rendus Mathématique", September 2015, n$^{o}$ 353, p. 1111-1115 [*DOI :* 10.1016/J.CRMA.2015.09.029], http://hal.upmc.fr/hal-01241300

[112] J. KEAN, S. RAINEY, M. MCFARLANE, C. DONALD, E. SCHNETTLER, A. KOHL, E. PONDEVILLE.*Fighting arbovirus transmission: natural and engineered control of vector competence in Aedes mosquitoes*, in "Insects", 2015, vol. 6, n$^{o}$ 1, p. 236–278

[113] I. C. KIM, B. PERTHAME, P. E. SOUGANIDIS.*Free boundary problems for tumor growth: a viscosity solutions approach*, in "Nonlinear Analysis: Theory, Methods and Applications", 2016, vol. 138, p. 207-228 [*DOI :* 10.1016/J.NA.2016.01.019], http://hal.upmc.fr/hal-01155696

[114] M. KOLWALCZYK, B. PERTHAME, N. VAUCHELET.*Transversal instability for the thermodiffusive reaction-diffusion system*, in "Chinese Annals of Mathematics - Series B", 2015, vol. 36, n$^{o}$ 5, p. 871-882, 13 pages, https://hal.archives-ouvertes.fr/hal-00939013

[115] T. LORENZI, R. H. CHISHOLM, J. CLAIRAMBAULT.*Tracking the evolution of cancer cell populations through the mathematical lens of phenotype-structured equations*, in "Biology Direct", December 2016, vol. 11, n$^{o}$ 1, 43 [*DOI :* 10.1186/S13062-016-0143-4], https://hal.inria.fr/hal-01355314

[116] T. LORENZI, R. H. CHISHOLM, L. DESVILLETTES, B. D. HUGHES.*Dissecting the dynamics of epigenetic changes in phenotype-structured populations exposed to fluctuating environments*, in "Journal of Theoretical Biology", September 2015, vol. 386, p. 166-176 [*DOI :* 10.1016/J.JTBI.2015.08.031], https://hal.inria.fr/hal-01237890

[117] T. LORENZI, R. H. CHISHOLM, A. LORZ.*Effects of an advection term in nonlocal Lotka-Volterra equations*, December 2015, https://hal.archives-ouvertes.fr/hal-01237529

[118] A. LORZ, T. LORENZI, J. CLAIRAMBAULT, A. ESCARGUEIL, B. PERTHAME.*Modeling the effects of space structure and combination therapies on phenotypic heterogeneity and drug resistance in solid tumors*, in "Bulletin of Mathematical Biology", January 2015, vol. 77, n$^{o}$ 1, p. 1-22 [*DOI :* 10.1007/s11538-014-0046-4], http://hal.upmc.fr/hal-00921266

[119] A. LORZ, T. LORENZI, M. E. HOCHBERG, J. CLAIRAMBAULT, B. PERTHAME.*Populational adaptive evolution, chemotherapeutic resistance and multiple anti-cancer therapies*, in "ESAIM: Mathematical Modelling and Numerical Analysis", March 2013, 23 [*DOI :* 10.1051/m2an/2012031], https://hal.archives-ouvertes.fr/hal-00714274

[120] A. MELLET, B. PERTHAME, F. QUIRÓS.*A Hele-Shaw Problem for Tumor Growth*, December 2015, working paper or preprint, http://hal.upmc.fr/hal-01241309

[121] L. A. MOREIRA, I. ITURBE-ORMAETXE, J. A. JEFFERY, G. LU, A. T. PYKE, L. M. HEDGES, B. C. ROCHA, S. HALL-MENDELIN, A. DAY, M. RIEGLER, L. E. HUGO, K. N. JOHNSON, B. H. KAY, E. A. MCGRAW, A. F. VAN DEN HURK, P. A. RYAN, S. L. O'NEILL.*A Wolbachia Symbiont in Aedes aegypti Limits Infection with Dengue, Chikungunya, and Plasmodium*, in "Cell", 2009, vol. 139, n$^{o}$ 7, p. 1268 - 1278

[122] A. OLIVIER.*Statistical analysis of growth-fragmentation models*, Université Paris Dauphine - Paris IX, November 2015, https://hal.archives-ouvertes.fr/tel-01235239

[123] A. OLIVIER.*How does variability in cells aging and growth rates influence the malthus parameter?*, in "Kinetic and Related Models ", June 2017, vol. 10, n$^{o}$ 2, p. 481-512 [*DOI :* 10.3934/krm.2017019], https://hal.archives-ouvertes.fr/hal-01274529

[124] K. PAKDAMAN, B. PERTHAME, D. SALORT.*Adaptation and Fatigue Model for Neuron Networks and Large Time Asymptotics in a Nonlinear Fragmentation Equation*, in "Journal of Mathematical Neuroscience", 2014, vol. 4, n$^{o}$ 1, 14 [*DOI :* 10.1186/2190-8567-4-14], https://hal.inria.fr/hal-01054561

[125] G. PEETERS, C. DEBBAUT, W. LALEMAN, A. FRIEBEL, D. MONBALIU, I. VANDER ELST, J. R. DETREZ, T. VANDECASTEELE, T. JOHANN, T. DE SCHRYVER, L. VAN HOOREBEKE, K. FAVERE, J. VERBEKE, D. DRASDO, S. HOEHME, P. SEGERS, P. CORNILLIE, W. H. DE VOS.*Corrigendum*, in "Journal of Anatomy", November 2017, vol. 231, n$^{o}$ 5, p. 786-786 [*DOI :* 10.1111/joa.12723], https://hal.inria.fr/hal-01968855

[126] B. PERTHAME.*Transport equations in biology*, Frontiers in Mathematics, Birkhäuser Verlag, Basel, 2007, x+198

[127] B. PERTHAME, F. QUIRÓS, M. TANG, N. VAUCHELET.*Derivation of a Hele-Shaw type system from a cell model with active motion*, July 2013, http://hal.upmc.fr/hal-00906168

[128] B. PERTHAME, F. QUIRÓS, J.-L. VÁZQUEZ.*The Hele-Shaw asymptotics for mechanical models of tumor growth*, in "Archive for Rational Mechanics and Analysis", 2014, vol. 212, p. 93-127 [*DOI :* 10.1007/s00205-013-0704-y], http://hal.upmc.fr/hal-00831932

[129] B. PERTHAME, D. SALORT.*On a voltage-conductance kinetic system for integrate and fire neural networks*, in "Kinetic and Related Models ", December 2013, vol. 6, n$^{o}$ 4, p. 841-864 [*DOI :* 10.3934/KRM.2013.6.841], http://hal.upmc.fr/hal-00871609

[130] B. PERTHAME, D. SALORT, G. WAINRIB.*Distributed synaptic weights in a LIF neural network and learning rules*, in "Physica D: Nonlinear Phenomena", 2017, vol. 353-354, p. 20-30 [*DOI :* 10.1016/J.PHYSD.2017.05.005], http://hal.upmc.fr/hal-01541093

[131] B. PERTHAME, M. TANG, N. VAUCHELET.*Traveling wave solution of the Hele-Shaw model of tumor growth with nutrient*, in "Mathematical Models and Methods in Applied Sciences", 2014, vol. 24, n$^{o}$ 13, p. 2601-2626, 25 pages, https://hal.archives-ouvertes.fr/hal-00931399

[132] B. PERTHAME, N. VAUCHELET.*Incompressible limit of mechanical model of tumor growth with viscosity*, in "Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences (1934–1990)", 2015, vol. 373, 20140283, 16 pages [*DOI :* 10.1098/RSTA.2014.0283], https://hal.archives-ouvertes.fr/hal-01066494

[133] B. PERTHAME, J. ZUBELLI.*On the inverse problem for a size-structured population model*, in "Inverse Problems", 2007, vol. 23, n$^{o}$ 3, p. 1037–1052

[134] D. PEURICHARD, E. AL.*Simple mechanical cues could explain adipose tissue morphology*, in "J. Theor Biol", 2017, https://doi.org/10.1016/j.jtbi.2017.06.030

[135] D. PEURICHARD, E. AL. *Extra-cellular matrix rigidity may dictate the fate of injury outcome*, in "J. Theor Biol (to appear)", 2019

[136] C. POUCHOL, J. CLAIRAMBAULT, A. LORZ, E. TRÉLAT.*Asymptotic analysis and optimal control of an integro-differential system modelling healthy and cancer cells exposed to chemotherapy*, December 2016, accepted in the J. Math. Pures et App., https://hal.archives-ouvertes.fr/hal-01416594

[137] S. PRIGENT, A. BALLESTA, F. CHARLES, N. LENUZZA, P. GABRIEL, L. M. TINE, H. REZAEI, M. DOUMIC.*An efficient kinetic model for assemblies of amyloid fibrils and its application to polyglutamine aggregation.*, in "PLoS ONE", 2012, vol. 7, n$^{o}$ 11, e43273 [*DOI :* 10.1371/JOURNAL.PONE.0043273], https://hal.archives-ouvertes.fr/hal-00778052

[138] L. ROBERT, M. HOFFMANN, N. KRELL, S. AYMERICH, J. ROBERT, M. DOUMIC.*Division in Escherichia coli is triggered by a size-sensing rather than a timing mechanism*, in "BMC Biology", 2014, vol. 12, n$^{o}$ 1, 17 [*DOI :* 10.1186/1741-7007-12-17], https://hal.inria.fr/hal-00981312

[139] F. SCHLIESS, S. HOEHME, S. HENKEL, A. GHALLAB, D. DRIESCH, J. BÖTTGER, R. GUTHKE, M. PFAFF, J. HENGSTLER, R. GEBHARDT, D. HÄUSSINGER, D. DRASDO, S. ZELLMER.*Integrated metabolic spatial-temporal model for the prediction of ammonia detoxification during liver damage and regeneration*, in "Hepatology", December 2014, vol. 60, n$^{o}$ 6, p. 2040–2051 [*DOI :* 10.1002/HEP.27136], https://hal.inria.fr/hal-01110646

[140] N. SFAKIANAKIS, D. PEURICHARD, A. BRUNK, C. SCHMEISER.*Modelling cell-cell collision and adhesion with the Filament Based Lamellipodium Model*, in "BIOMATH", 2019, http://dx.doi.org/10.11145/j.biomath.2018.11.097

[141] S. P. SINKINS.*Wolbachia and cytoplasmic incompatibility in mosquitoes*, in "Insect Biochemistry and Molecular Biology", 2004, vol. 34, n$^{o}$ 7, p. 723 - 729, Molecular and population biology of mosquitoes

[142] P. VAN LIEDEKERKE, J. NEITSCH, T. JOHANN, K. ALESSANDRI, P. NASSOY, D. DRASDO.*Quantitative modeling identifies robust predictable stress response of growing CT26 tumor spheroids under variable conditions*, December 2016, working paper or preprint, https://hal.inria.fr/hal-01421179

[143] P. VAN LIEDEKERKE, M. M. PALM, N. JAGIELLA, D. DRASDO.*Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results*, in "Computational Particle Mechanics", Nov 2015, vol. 2, n$^{o}$ 4, p. 401–444, http://dx.doi.org/10.1007/s40571-015-0082-3

[144] T. WALKER, S. P. SINKINS.*Biological control of arbovirus vectors*, in "Arboviruses: Molecular Biology, Evolution and Control. Caister Academic Press, Norfolk, UK", 2016, p. 291–302

[145] Y. YIN, O. SEDLACZEK, B. MÜLLER, A. WARTH, M. GONZÁLEZ-VALLINAS, B. LAHRMANN, N. GRABE, H.-U. KAUCZOR, K. BREUHAHN, I. VIGNON-CLEMENTEL, D. DRASDO.*Tumor cell load and heterogeneity estimation from diffusion-weighted MRI calibrated with histological data: an example from lung cancer*, in "IEEE Transactions on Medical Imaging", 2017 [*DOI :* 10.1109/TMI.2017.2698525], https://hal.inria.fr/hal-01421398

# Project-Team MATHERIALS

## MATHematics for MatERIALS

IN COLLABORATION WITH: Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique (CERMICS)

IN PARTNERSHIP WITH:
**Ecole des Ponts ParisTech**

RESEARCH CENTER
**Paris**

THEME
**Numerical schemes and simulations**

# Table of contents

# Project-Team MATHERIALS

*Creation of the Team: 2014 January 01, updated into Project-Team: 2015 April 01*

**Keywords:**

**Computer Science and Digital Science:**

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.2.4. - Statistical methods
A6.2.7. - High performance computing
A6.3.1. - Inverse problems
A6.3.4. - Model reduction
A6.4.1. - Deterministic control

**Other Research Topics and Application Domains:**

B1.1.2. - Molecular and cellular biology
B4.3.4. - Solar Energy
B5.3. - Nanotechnology
B5.5. - Materials
B9.5.2. - Mathematics
B9.5.3. - Physics
B9.5.4. - Chemistry

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Claude Le Bris [Team leader, Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Sébastien Boyaval [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Éric Cancès [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Virginie Ehrlacher [Ecole Nationale des Ponts et Chaussées, Researcher]
Frédéric Legoll [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Tony Lelièvre [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Antoine Levitt [Inria, Researcher]
Gabriel Stoltz [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]

**PhD Students**

Robert Benda [Ecole Polytechnique, from Sep 2018]
Grégoire Ferré [Ecole Nationale des Ponts et Chaussées]
Marc Josien [Ministère de la transition écologique et solidaire, until Aug 2018]
Adrien Lesage [Ecole Nationale des Ponts et Chaussées]
Mouad Ramil [Ecole Nationale des Ponts et Chaussées]
Pierre-Loïk Rothé [Ecole Nationale des Ponts et Chaussées]

Julien Roussel [Ecole Nationale des Ponts et Chaussées, until Aug 2018]
Laura Silva Lopes [Ecole Nationale des Ponts et Chaussées]
Sami Siraj-Dine [Univ Paris-Est Marne La Vallée]
Pierre Terrier [Ministère de la transition écologique et solidaire, until Sep 2018]

**Post-Doctoral Fellows**

Florent Hédin [Ecole Nationale des Ponts et Chaussées, until Nov 2018]
Dena Kazerani [Inria, until Sep 2018]

**Administrative Assistants**

Sabrine Boumizy [Inria, until Feb 2018]
Hélène Milome [Inria, from Mar 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

The MATHERIALS project-team has been created jointly by the École des Ponts ParisTech (ENPC) and Inria in 2015. It is the follow-up and an extension of the former project-team MICMAC originally created in October 2002. It is hosted by the CERMICS laboratory (Centre d'Enseignement et de Recherches en Mathématiques et Calcul Scientifique) at École des Ponts. The permanent research scientists of the project-team have positions at CERMICS and at two other laboratories of École des Ponts: Institut Navier and Laboratoire Saint-Venant. The scientific focus of the project-team is to analyze and improve the numerical schemes used in the simulation of computational chemistry at the microscopic level and to create simulations coupling this microscopic scale with meso- or macroscopic scales (possibly using parallel algorithms). Over the years, the project-team has accumulated an increasingly solid expertise on such topics, which are traditionally not well known by the community in applied mathematics and scientific computing. One of the major achievements of the project-team is to have created a corpus of literature, authoring books and research monographs on the subject [1], [2], [3], [5], [6] that other scientists may consult in order to enter the field.

# 3. Research Program

## 3.1. Research Program

Our group, originally only involved in electronic structure computations, continues to focus on many numerical issues in quantum chemistry, but now expands its expertise to cover several related problems at larger scales, such as molecular dynamics problems and multiscale problems. The mathematical derivation of continuum energies from quantum chemistry models is one instance of a long-term theoretical endeavour.

### *3.1.1. Electronic structure of large systems*

Quantum Chemistry aims at understanding the properties of matter through the modelling of its behavior at a subatomic scale, where matter is described as an assembly of nuclei and electrons. At this scale, the equation that rules the interactions between these constitutive elements is the Schrödinger equation. It can be considered (except in few special cases notably those involving relativistic phenomena or nuclear reactions) as a universal model for at least three reasons. First it contains all the physical information of the system under consideration so that any of the properties of this system can in theory be deduced from the Schrödinger equation associated to it. Second, the Schrödinger equation does not involve any empirical parameters, except some fundamental constants of Physics (the Planck constant, the mass and charge of the electron, ...); it can thus be written for any kind of molecular system provided its chemical composition, in terms of natures of nuclei and number of electrons, is known. Third, this model enjoys remarkable predictive capabilities, as confirmed by comparisons with a large amount of experimental data of various types. On the other hand, using this high quality model requires working with space and time scales which are both very tiny: the typical size of the electronic cloud of

an isolated atom is the Angström ($10^{-10}$ meters), and the size of the nucleus embedded in it is $10^{-15}$ meters; the typical vibration period of a molecular bond is the femtosecond ($10^{-15}$ seconds), and the characteristic relaxation time for an electron is $10^{-18}$ seconds. Consequently, Quantum Chemistry calculations concern very short time (say $10^{-12}$ seconds) behaviors of very small size (say $10^{-27}$ m$^3$) systems. The underlying question is therefore whether information on phenomena at these scales is useful in understanding or, better, predicting macroscopic properties of matter. It is certainly not true that *all* macroscopic properties can be simply upscaled from the consideration of the short time behavior of a tiny sample of matter. Many of them derive from ensemble or bulk effects, that are far from being easy to understand and to model. Striking examples are found in solid state materials or biological systems. Cleavage, the ability of minerals to naturally split along crystal surfaces (e.g. mica yields to thin flakes), is an ensemble effect. Protein folding is also an ensemble effect that originates from the presence of the surrounding medium; it is responsible for peculiar properties (e.g. unexpected acidity of some reactive site enhanced by special interactions) upon which vital processes are based. However, it is undoubtedly true that *many* macroscopic phenomena originate from elementary processes which take place at the atomic scale. Let us mention for instance the fact that the elastic constants of a perfect crystal or the color of a chemical compound (which is related to the wavelengths absorbed or emitted during optic transitions between electronic levels) can be evaluated by atomic scale calculations. In the same fashion, the lubricative properties of graphite are essentially due to a phenomenon which can be entirely modeled at the atomic scale. It is therefore reasonable to simulate the behavior of matter at the atomic scale in order to understand what is going on at the macroscopic one. The journey is however a long one. Starting from the basic principles of Quantum Mechanics to model the matter at the subatomic scale, one finally uses statistical mechanics to reach the macroscopic scale. It is often necessary to rely on intermediate steps to deal with phenomena which take place on various *mesoscales*. It may then be possible to couple one description of the system with some others within the so-called *multiscale* models. The sequel indicates how this journey can be completed focusing on the first smallest scales (the subatomic one), rather than on the larger ones. It has already been mentioned that at the subatomic scale, the behavior of nuclei and electrons is governed by the Schrödinger equation, either in its time-dependent form or in its time-independent form. Let us only mention at this point that

- both equations involve the quantum Hamiltonian of the molecular system under consideration; from a mathematical viewpoint, it is a self-adjoint operator on some Hilbert space; *both* the Hilbert space and the Hamiltonian operator depend on the nature of the system;
- also present into these equations is the wavefunction of the system; it completely describes its state; its $L^2$ norm is set to one.

The time-dependent equation is a first-order linear evolution equation, whereas the time-independent equation is a linear eigenvalue equation. For the reader more familiar with numerical analysis than with quantum mechanics, the linear nature of the problems stated above may look auspicious. What makes the numerical simulation of these equations extremely difficult is essentially the huge size of the Hilbert space: indeed, this space is roughly some symmetry-constrained subspace of $L^2(\mathbb{R}^d)$, with $d = 3(M + N)$, $M$ and $N$ respectively denoting the number of nuclei and the number of electrons the system is made of. The parameter $d$ is already 39 for a single water molecule and rapidly reaches $10^6$ for polymers or biological molecules. In addition, a consequence of the universality of the model is that one has to deal at the same time with several energy scales. In molecular systems, the basic elementary interaction between nuclei and electrons (the two-body Coulomb interaction) appears in various complex physical and chemical phenomena whose characteristic energies cover several orders of magnitude: the binding energy of core electrons in heavy atoms is $10^4$ times as large as a typical covalent bond energy, which is itself around 20 times as large as the energy of a hydrogen bond. High precision or at least controlled error cancellations are thus required to reach chemical accuracy when starting from the Schrödinger equation. Clever approximations of the Schrödinger problems are therefore needed. The main two approximation strategies, namely the Born-Oppenheimer-Hartree-Fock and the Born-Oppenheimer-Kohn-Sham strategies, end up with large systems of coupled *nonlinear* partial differential equations, each of these equations being posed on $L^2(\mathbb{R}^3)$. The size of the underlying functional space is thus reduced at the cost of a dramatic increase of the mathematical complexity of the problem: nonlinearity. The mathematical and numerical analysis of the resulting models has been the major concern

of the project-team for a long time. In the recent years, while part of the activity still follows this path, the focus has progressively shifted to problems at other scales.

As the size of the systems one wants to study increases, more efficient numerical techniques need to be resorted to. In computational chemistry, the typical scaling law for the complexity of computations with respect to the size of the system under study is $N^3$, $N$ being for instance the number of electrons. The Holy Grail in this respect is to reach a linear scaling, so as to make possible simulations of systems of practical interest in biology or material science. Efforts in this direction must address a large variety of questions such as

- how can one improve the nonlinear iterations that are the basis of any *ab initio* models for computational chemistry?
- how can one more efficiently solve the inner loop which most often consists in the solution procedure for the linear problem (with frozen nonlinearity)?
- how can one design a sufficiently small variational space, whose dimension is kept limited while the size of the system increases?

An alternative strategy to reduce the complexity of *ab initio* computations is to try to couple different models at different scales. Such a mixed strategy can be either a sequential one or a parallel one, in the sense that

- in the former, the results of the model at the lower scale are simply used to evaluate some parameters that are inserted in the model for the larger scale: one example is the parameterized classical molecular dynamics, which makes use of force fields that are fitted to calculations at the quantum level;
- while in the latter, the model at the lower scale is concurrently coupled to the model at the larger scale: an instance of such a strategy is the so called QM/MM coupling (standing for Quantum Mechanics/Molecular Mechanics coupling) where some part of the system (typically the reactive site of a protein) is modeled with quantum models, that therefore accounts for the change in the electronic structure and for the modification of chemical bonds, while the rest of the system (typically the inert part of a protein) is coarse grained and more crudely modeled by classical mechanics.

The coupling of different scales can even go up to the macroscopic scale, with methods that couple a microscopic representation of matter, or at least a mesoscopic one, with the equations of continuum mechanics at the macroscopic level.

### *3.1.2. Computational Statistical Mechanics*

The orders of magnitude used in the microscopic representation of matter are far from the orders of magnitude of the macroscopic quantities we are used to: The number of particles under consideration in a macroscopic sample of material is of the order of the Avogadro number $\mathcal{N}_A \sim 6 \times 10^{23}$, the typical distances are expressed in Å ($10^{-10}$ m), the energies are of the order of $k_{\mathrm{B}}T \simeq 4 \times 10^{-21}$ J at room temperature, and the typical times are of the order of $10^{-15}$ s.

To give some insight into such a large number of particles contained in a macroscopic sample, it is helpful to compute the number of moles of water on earth. Recall that one mole of water corresponds to 18 mL, so that a standard glass of water contains roughly 10 moles, and a typical bathtub contains $10^5$ mol. On the other hand, there are approximately $10^{18}$ m$^3$ of water in the oceans, *i.e.*$7 \times 10^{22}$ mol, a number comparable to the Avogadro number. This means that inferring the macroscopic behavior of physical systems described at the microscopic level by the dynamics of several millions of particles only is like inferring the ocean's dynamics from hydrodynamics in a bathtub...

For practical numerical computations of matter at the microscopic level, following the dynamics of every atom would require simulating $\mathcal{N}_A$ atoms and performing $\mathrm{O}(10^{15})$ time integration steps, which is of course impossible! These numbers should be compared with the current orders of magnitude of the problems that can be tackled with classical molecular simulation, where several millions of atoms only can be followed over time scales of the order of a few microseconds.

Describing the macroscopic behavior of matter knowing its microscopic description therefore seems out of reach. Statistical physics allows us to bridge the gap between microscopic and macroscopic descriptions of matter, at least on a conceptual level. The question is whether the estimated quantities for a system of $N$ particles correctly approximate the macroscopic property, formally obtained in the thermodynamic limit $N \to +\infty$ (the density being kept fixed). In some cases, in particular for simple homogeneous systems, the macroscopic behavior is well approximated from small-scale simulations. However, the convergence of the estimated quantities as a function of the number of particles involved in the simulation should be checked in all cases.

Despite its intrinsic limitations on spatial and timescales, molecular simulation has been used and developed over the past 50 years, and its number of users keeps increasing. As we understand it, it has two major aims nowadays.

First, it can be used as a *numerical microscope*, which allows us to perform "computer" experiments. This was the initial motivation for simulations at the microscopic level: physical theories were tested on computers. This use of molecular simulation is particularly clear in its historic development, which was triggered and sustained by the physics of simple liquids. Indeed, there was no good analytical theory for these systems, and the observation of computer trajectories was very helpful to guide the physicists' intuition about what was happening in the system, for instance the mechanisms leading to molecular diffusion. In particular, the pioneering works on Monte-Carlo methods by Metropolis *et al.*, and the first molecular dynamics simulation of Alder and Wainwright were performed because of such motivations. Today, understanding the behavior of matter at the microscopic level can still be difficult from an experimental viewpoint (because of the high resolution required, both in time and in space), or because we simply do not know what to look for! Numerical simulations are then a valuable tool to test some ideas or obtain some data to process and analyze in order to help assessing experimental setups. This is particularly true for current nanoscale systems.

Another major aim of molecular simulation, maybe even more important than the previous one, is to compute macroscopic quantities or thermodynamic properties, typically through averages of some functionals of the system. In this case, molecular simulation is a way to obtain *quantitative* information on a system, instead of resorting to approximate theories, constructed for simplified models, and giving only qualitative answers. Sometimes, these properties are accessible through experiments, but in some cases only numerical computations are possible since experiments may be unfeasible or too costly (for instance, when high pressure or large temperature regimes are considered, or when studying materials not yet synthesized). More generally, molecular simulation is a tool to explore the links between the microscopic and macroscopic properties of a material, allowing one to address modelling questions such as "Which microscopic ingredients are necessary (and which are not) to observe a given macroscopic behavior?"

### *3.1.3. Homogenization and related problems*

Over the years, the project-team has developed an increasing expertise on how to couple models written at the atomistic scale with more macroscopic models, and, more generally, an expertise in multiscale modelling for materials science.

The following observation motivates the idea of coupling atomistic and continuum representation of materials. In many situations of interest (crack propagation, presence of defects in the atomistic lattice, ...), using a model based on continuum mechanics is difficult. Indeed, such a model is based on a macroscopic constitutive law, the derivation of which requires a deep qualitative and quantitative understanding of the physical and mechanical properties of the solid under consideration. For many solids, reaching such an understanding is a challenge, as loads they are subjected to become larger and more diverse, and as experimental observations helping designing such models are not always possible (think of materials used in the nuclear industry). Using an atomistic model in the whole domain is not possible either, due to its prohibitive computational cost. Recall indeed that a macroscopic sample of matter contains a number of atoms on the order of $10^{23}$. However, it turns out that, in many situations of interest, the deformation that we are looking for is not smooth in *only a small part* of the solid. So, a natural idea is to try to take advantage of both models, the continuum mechanics one and the atomistic one, and to couple them, in a domain decomposition spirit. In most of the domain, the

deformation is expected to be smooth, and reliable continuum mechanics models are then available. In the rest of the domain, the expected deformation is singular, so that one needs an atomistic model to describe it properly, the cost of which remains however limited as this region is small.

From a mathematical viewpoint, the question is to couple a discrete model with a model described by PDEs. This raises many questions, both from the theoretical and numerical viewpoints:

- first, one needs to derive, from an atomistic model, continuum mechanics models, under some regularity assumptions that encode the fact that the situation is smooth enough for such a macroscopic model to provide a good description of the materials;
- second, couple these two models, e.g. in a domain decomposition spirit, with the specificity that models in both domains are written in a different language, that there is no natural way to write boundary conditions coupling these two models, and that one would like the decomposition to be self-adaptive.

More generally, the presence of numerous length scales in material science problems represents a challenge for numerical simulation, especially when some *randomness* is assumed on the materials. It can take various forms, and includes defects in crystals, thermal fluctuations, and impurities or heterogeneities in continuous media. Standard methods available in the literature to handle such problems often lead to very costly computations. Our goal is to develop numerical methods that are more affordable. Because we cannot embrace all difficulties at once, we focus on a simple case, where the fine scale and the coarse-scale models can be written similarly, in the form of a simple elliptic partial differential equation in divergence form. The fine scale model includes heterogeneities at a small scale, a situation which is formalized by the fact that the coefficients in the fine scale model vary on a small length scale. After homogenization, this model yields an effective, macroscopic model, which includes no small scale. In many cases, a sound theoretical groundwork exists for such homogenization results. The difficulty stems from the fact that the models generally lead to prohibitively costly computations. For such a case, simple from the theoretical viewpoint, our aim is to focus on different practical computational approaches to speed-up the computations. One possibility, among others, is to look for specific random materials, relevant from the practical viewpoint, and for which a dedicated approach can be proposed, that is less expensive than the general approach.

# 4. Highlights of the Year

## 4.1. Highlights of the Year

### *4.1.1. Awards*

Claude Le Bris was selected to deliver the Coxeter lectures at the Fields Institute in Toronto and the Aziz lectures at the University of Maryland.

Florent Hédin received the “Best student/postdoc oral presentation” award at the $7^{\text{th}}$ Workshop on Parallel-in-Time methods, Roscoff, France, May.

# 5. New Software and Platforms

## 5.1. simol

KEYWORDS: Molecular simulation - Quantum chemistry - Statistical physics - C++ - OpenMP
FUNCTIONAL DESCRIPTION: Molecular simulation software written in C++

- Contact: Gabriel Stoltz

# 6. New Results

## 6.1. Electronic structure calculations

**Participants:** Robert Benda, Éric Cancès, Virginie Ehrlacher, Antoine Levitt, Sami Siraj-Dine, Gabriel Stoltz.

In electronic structure calculation as in most of our scientific endeavors, we pursue a twofold goal: placing the models on a sound mathematical grounding by an appropriate mathematical analysis, and improving the numerical approaches by a dedicated numerical analysis.

### *6.1.1. Mathematical analysis*

The members of the team have continued their systematic study of the properties of materials in the reduced Hartree-Fock approximation, a model striking a good balance between mathematical tractability and the ability to reproduce qualitatively complex effects.

E. Cancès and G. Stoltz have studied with L. Cao models for certain extended defects in materials [37]. These extended defects typically correspond to taking out a slab of finite width in the three-dimensional homogeneous electron gas. The work is performed in the framework of the reduced Hartree-Fock model with either Yukawa or Coulomb interactions, using techniques previously developed to study local perturbations of the free-electron gas. It is shown that the model admits minimizers, and that Yukawa ground state energies and density matrices converge to ground state Coulomb energies and density matrices as the Yukawa parameter tends to zero. These minimizers are unique for Yukawa interactions, and are characterized by a self-consistent equation. Numerical simulations show evidence of Friedel oscillations in the total electronic density.

A. Levitt has examined the phenomenon of screening in materials. In [54] he has studied the effect of adding a small charge to a periodic system modeled by the reduced Hartree-Fock at finite temperature. He has showed that the reaction potential created by the rearrangement of the electrons counteracts exactly the free charge, so that the effective interaction in such systems is short-range. The proof proceeds by studying the properties of the linear response operator, which also sheds some light on the charge-sloshing instability seen in numerical methods to solve the self-consistent equations.

### *6.1.2. Numerical analysis*

E. Cancès has pursued his long-term collaboration with Y. Maday (Sorbonne Université) on the numerical analysis of linear and nonlinear eigenvalue problems. Together with G. Dusson (Warwick, United Kingdom), B. Stamm (Aachen, Germany), and M. Vohralík (Inria SERENA), they have designed *a posteriori* error estimates for conforming numerical approximations of the Laplace eigenvalue problem with homogeneous Dirichlet boundary conditions. In [38], they prove *a priori* error estimates for the perturbation-based post-processing of the plane-wave approximation of Schrödinger equations introduced and tested numerically in previous works. They consider a Schrödinger operator $H = -\frac{1}{2}\Delta + V$ on $L^2(\Omega)$, where $\Omega$ is a cubic box with periodic boundary conditions. The quantities of interest are, on the one hand, the ground-state energy defined as the sum of the lowest $N$ eigenvalues of $H$ , and, on the other hand, the ground-state density matrix, that is the spectral projector on the vector space spanned by the associated eigenvectors. Such a problem is central in first-principle molecular simulation, since it corresponds to the so-called linear subproblem in Kohn-Sham density functional theory (DFT). Interpreting the exact eigenpairs of H as perturbations of the numerical eigenpairs obtained by a variational approximation in a plane-wave (i.e. Fourier) basis, they compute first-order corrections for the eigenfunctions, which are turned into corrections on the ground-state density matrix. This allows them to increase the accuracy of both the ground-state energy and the ground-state density matrix at a low computational extra-cost. Indeed, the computation of the corrections only requires the computation of the residual of the solution in a larger plane-wave basis and $2N$ Fast Fourier Transforms.

Implicit solvation models aim at computing the properties of a molecule in solution (most chemical reactions take place in the liquid phase) by replacing all the solvent molecules but the few ones strongly interacting with the solute, by an effective continuous medium accounting for long-range electrostatics. E. Cancès, Y. Maday (Sorbonne Université), and B. Stamm (Aachen, Germany) have introduced a few years ago a very efficient domain decomposition method for the simulation of large molecules in the framework of the so-called COSMO implicit solvation models. In collaboration with F. Lipparini and B. Mennucci (Chemistry, Pisa, Italy) and J.-P. Piquemal (Sorbonne Université), they have implemented this algorithm in widely used computational software products (Gaussian and Tinker). Together with L. Lagardère (Sorbonne Université) and G. Scalmani (Gaussian Inc., USA), they illustrate in [29] the domain decomposition COSMO (ddCOSMO) implementation and how to couple it with an existing classical or quantum mechanical (QM) codes. They review in detail what input needs to be provided to ddCOSMO and how to assemble it, describe how the ddCOSMO equations are solved and how to process the results in order to assemble the required quantities, such as Fock matrix contributions for the QM case, or forces for the classical one. Throughout the paper, they make explicit references to the ddCOSMO module, which is an open source, Fortran 90 implementation of ddCOSMO that can be downloaded and distributed under the LGPL license.

E. Cancès, V. Ehrlacher and A. Levitt, together with D. Gontier (Dauphine) and D. Lombardi (Inria REO), have studied the convergence of properties of periodic systems as the size of the computing domain is increased. This convergence is known to be difficult in the case of metals. They have characterized in [39] the speed of convergence for a number of schemes in the metallic case, and have studied the properties of a widely used numerical method that adds an artificial electronic temperature.

A. Levitt has continued his study of Wannier functions in periodic systems. With A. Damle (Cornell, USA) and L. Lin (Berkeley, USA), they have proposed an efficient numerical method for the computation of maximally-localized Wannier functions in metals, and have showed on the example of the free electron gas that they are not in general exponentially localized [42]. With D. Gontier (Dauphine) and S. Siraj-Dine, they proposed a new method for the computation of Wannier functions which applies to any insulator, and in particular to the difficult case of topological insulators [45].

## 6.2. Computational Statistical Physics

**Participants:** Grégoire Ferré, Florent Hédin, Frédéric Legoll, Tony Lelièvre, Mouad Ramil, Julien Roussel, Laura Silva Lopes, Gabriel Stoltz, Pierre Terrier.

The objective of computational statistical physics is to compute macroscopic properties of materials starting from a microscopic description, using concepts of statistical physics (thermodynamic ensembles and molecular dynamics). The contributions of the team can be divided into four main topics: (i) the development of methods for sampling the configuration space; (ii) the numerical analysis of such methods; (iii) the efficient computation of dynamical properties which requires to sample metastable trajectories; (iv) coarse-graining techniques to reduce the computational cost of molecular dynamic simulations and gain some insights on the models.

### *6.2.1. Sampling of the configuration space: new algorithms and applications*

New numerical methods in order to sample probability measures on the configuration space have been developed: either measures supported on submanifolds, or stationary states of stochastic dynamics. First, in [51], T. Lelièvre and G. Stoltz, together with M. Rousset (Inria Rennes, France) have studied how to sample probability measures supported on submanifolds, by adding an extra momentum variable to the state of the system, and discretizing the associated Hamiltonian dynamics with some stochastic perturbation in the extra variable. In order to avoid biases in the invariant probability measures sampled by discretizations of these stochastically perturbed Hamiltonian dynamics, a Metropolis rejection procedure can be considered. The so-obtained scheme belongs to the class of generalized Hybrid Monte Carlo (GHMC) algorithms. However, the usual method has to be generalized using a procedure suggested by Goodman, Holmes-Cerfon and Zappa for Metropolis random walks on submanifolds, where a reverse projection check is performed to enforce the reversibility of the algorithm for large timesteps and hence avoid biases in the invariant measure. A full mathematical analysis of such procedures is provided, as well as numerical experiments demonstrating the

importance of the reverse projection check on simple toy examples. Second, the work [55] by J. Roussel and G. Stoltz focuses on the use of control variates for non-equilibrium systems. Whereas most variance reduction methods rely on the knowledge of the invariant probability measure, this latter is not explicit out of equilibrium. Control variates offer an attractive alternative in this framework. J. Roussel and G. Stoltz have proposed a general strategy for constructing an efficient control variate, relying on physical simplifications of the dynamics. The authors provide an asymptotic analysis of the variance reduction in a perturbative framework, along with extensive numerical tests on three different systems.

In terms of applications of such sampling techniques, members of the project-team have been working on two different subjects: random matrices models and adaptive techniques to compute large deviation rate functionals. The paper [16] was written by G. Ferré and D. Chafaï (Université Paris Dauphine, France), following the simple idea: the eigenvalues of random matrices are distributed according to Boltzmann–Gibbs measures, but researchers in this field do not use techniques from statistical physics for numerical investigations. The authors therefore used a Hamiltonian Monte Carlo algorithm to investigate numerically conjectures about random matrices and related Coulomb gases. The next step is to add constraints to these systems to understand better the behavior of random matrices with constraints and the large size limit of their spectra (the algorithm mentioned above to sample probability measures supported on submanifolds may be useful in this context). The work [19] focuses on computing free energies and entropy functions, as they arise in large deviations theory, through adaptive techniques. It is actually in the spirit of techniques used in mathematical finance, adapted to the statistical mechanics context, and enriched with new estimators based on variational representations of entropy functions. These tools have been pioneered by H. Touchette (Stellenbosch University, South Africa), with whom the paper was written by G. Ferré.

### *6.2.2. Sampling of the configuration space: numerical analysis*

Concerning the numerical analysis of sampling techniques of probability measures on the configuration space, let us mention three works.

First, in [44], G. Ferré and G. Stoltz study the numerical errors that arise when a stochastic differential equation (SDE) is discretized in order to compute scaled cumulant functions (or free energy) and ergodic properties of Feynman–Kac semigroups. These quantities naturally arise in large deviations theory, for estimating probabilities of rare events. This analysis is made difficult by the nonlinear (mean field) feature of the dynamics at hand. The obtained estimates generalize previous results on the numerical analysis of ergodic properties of discretized SDEs. As a theoretical extension of the previous work, the purpose of the work [43] by G. Ferré and G. Stoltz, in collaboration with M. Rousset (Inria Rennes, France), is to provide further theoretical investigations on the long time behavior of Feynman–Kac semigroups. More precisely, it aims at giving practical criteria for these nonlinear semigroups to have a limit, and makes precise in which sense this limit is to be understood. This was an open problem so far for systems evolving in unbounded configuration spaces, which was addressed through Lyapunov function techniques. Although theoretical, these results are of practical importance since, if these dynamics do not have a well-defined long time behavior, it is hopeless to try to compute rare events.

Finally, together with C. Andrieu (Univ. Bristol, United-Kingdom), A. Durmus (ENS Saclay, France) and N. Nüsken (Univ. Potsdam, Germany), J. Roussel derived in [32] spectral gap estimates for several Piecewise Deterministic Markov Processes (PDMPs), namely the Randomized Hamiltonian Monte Carlo, the Zig-Zag process and the Bouncy Particle Sampler. The hypocoercivity technique provides estimates with explicit dependence on the parameters of the dynamics. Moreover the general framework considered allows to compare quantitatively the bounds found for the different methods. Such PDMDs are currently more and more used as efficient sampling tools, but their theoretical properties are still not yet well understood.

### *6.2.3. Sampling of dynamical properties and rare events*

The sampling of dynamical properties along molecular dynamics trajectories is crucial to get access to important quantities such as transition rates or reactive paths. This is difficult numerically because of the metastability of trajectories. Members of the project-team are following two numerical approaches to sample

metastable trajectories: the accelerated dynamics *à la* A.F. Voter and the adaptive multilevel splitting (AMS) technique to sample reactive paths between metastable states.

Concerning the mathematical analysis of the accelerated dynamics, in [50], T. Lelièvre reviews the recent mathematical approaches to justify these numerical methods, using the notion of quasi-stationary distribution. Moreover, in [49], T. Lelièvre together with D. Le Peutrec (Université de Paris Saclay, France) and G. Di Gesu and B. Nectoux (TU Wien, Austria) give an overview of the results obtained during the PhD of B. Nectoux. Using the quasi-stationary distribution approach and tools from semi-classical analysis, one can justify the use of kinetic Monte Carlo models parametrized by the Eyring-Kramers formulas to describe exit events from metastable states, for the overdamped Langevin dynamics. Concerning the implementation, in [22], F. Hédin and T. Lelièvre test the Generalized Parallel Replica algorithm to biological systems, and obtain strong linear scalability, providing up to 70% of the maximum possible speedup on several hundreds of CPUs. The "Parallel Replica" (ParRep) dynamics is known for allowing to simulate very long trajectories of metastable Langevin dynamics in the materials science community, but it relies on assumptions that can hardly be transposed to the world of biochemical simulations. The later developed "Generalized ParRep" variant solves those issues, and it had not been applied to significant systems of interest so far. Finally, let us mention the work [27] where T. Lelièvre together with J. Reygner (Ecole des Ponts, France) and L. Pillaud-Vivien (Inria Paris, France) analyze mathematically the Fleming-Viot particle process in the simple case of a finite state space. This Fleming-Viot particle process is a key ingredient of the Generalized ParRep algorithm mentioned above, in order to both approximate the convergence time to the quasi-stationary distribution, and to efficiently sample it.

Concerning the AMS technique, in [36], T. Lelièvre and C.-E. Bréhier (ENS Lyon, France) test new importance functions to compute rare events associated with the law of the solution to a stochastic differential equation at a given fixed time. This can be used for example to estimate the rate functional for large deviation principle applied to time averages.

### *6.2.4. Coarse-graining*

In two related works, members of the project-team have studied the quality of the effective dynamics derived from a high dimension stochastic differential equation on a few degrees of freedom, using a projection approach *à la Mori-Zwanzig*. More precisely, in [48], F. Legoll, T. Lelièvre and U. Sharma obtain precise error bounds in the case of non reversible dynamics. This analysis also aims at discussing what is a good notion of mean force for non reversible systems. In [53], T. Lelièvre together with W. Zhang (ZIB, Germany) extend previous results on pathwise error estimates for such effective dynamics to the case of nonlinear vectorial reaction coordinates.

Once a good coarse-grained model has been obtained, one can try to use it in order to get a better integrator of the original dynamic in the spirit of a predictor-corrector method. In [52], T. Lelièvre together with G. Samaey and P. Zielinski (KU Leuven, Belgium) analyze such a micro-macro acceleration method for the Monte Carlo simulation of stochastic differential equations with time-scale separation between the (fast) evolution of individual trajectories and the (slow) evolution of the macroscopic function of interest.

## 6.3. Homogenization

**Participants:** Virginie Ehrlacher, Marc Josien, Claude Le Bris, Frédéric Legoll, Adrien Lesage, Pierre-Loïk Rothé.

### *6.3.1. Deterministic non-periodic systems*

In homogenization theory, members of the project-team have pursued their ongoing systematic study of perturbations of periodic problems (by local and nonlocal defects). This has been done in two different directions.

For linear elliptic equations, C. Le Bris has written, in collaboration with X. Blanc (Paris Diderot, France) and P.-L. Lions (Collège de France, France), two manuscripts that present a more versatile proof of the existence of a corrector function for periodic problems with local defects, and also extend the results: the first manuscript [34] addresses the case of an equation (or a system) in divergence form, while the second manuscript [12] extends the analysis to advection-diffusion equations.

Second, they have also provided more details on the quality of approximation achieved by their theory. The fact that a corrector exists with suitable properties allows one to quantify the rate of convergence of the two-scale expansion using that corrector to the actual exact solution, as the small homogenization parameter $\varepsilon$ vanishes. These works by C. Le Bris, in collaboration with X. Blanc and M. Josien (and in the context of the PhD thesis of the latter), will be presented in a series of manuscripts in preparation. The precise results have been announced in [11] and proven in [33]. A related study [47] has been performed by M. Josien and addresses issues regarding periodic Green functions.

Also in the context of homogenization theory, C. Le Bris and F. Legoll have initiated a collaboration with R. Cottereau (Ecole Centrale and now CNRS Marseille, France). The topic is in some sense a follow-up on both an earlier work of R. Cottereau and the series of works completed by C. Le Bris and F. Legoll in collaboration with K. Li and next S. Lemaire over the years. Schematically, the purpose of the work is to determine the homogenized coefficient for a medium without explicitly performing a homogenization approach nor using a MsFEM type approach. In earlier works, an approximation approach, somewhat engineering-style, was designed. The purpose now is to examine the performance of this approach in the context of the so-called Arlequin method, a very popular method in the mechanical engineering community. One couples a sub-region of the medium where a homogeneous model is employed, along with a complementary sub-region where the original multiscale model is solved explicitly. The coupling is performed using the Arlequin method. Then, one optimizes a suitable criterion so that optimization leads to an homogeneous sub-region indeed described by the homogenized coefficient seeked for. Some numerical analysis questions, together with practical perspectives for computational enhancements of the approach, are currently examined.

Finally, C. Le Bris has informally participated into the supervision of the master thesis of S. Wolf (Ecole Normale Supérieure, Paris, France), and in this context performed some works in interaction with the student and X. Blanc. The purpose is to investigate perturbations of periodic homogenization problems when the perturbation is geometric in nature. The test case considered is that of a domain perforated by holes the locations of which are not necessarily periodic, but only periodic up to a local perturbation. The results proven, on the prototypical Poisson equation, are natural extensions of the celebrated results by J.-L. Lions published in the late 1960s for the periodic case. This provides a proof of concept, showing that perturbations of a periodic geometry are also possible, a fact that will be more thoroughly investigated in the near future within the above mentioned collaboration.

### *6.3.2. Stochastic homogenization*

The project-team has pursued its efforts in the field of stochastic homogenization of elliptic equations, aiming at designing numerical approaches that are practically relevant and keep the computational workload limited.

Using standard homogenization theory, one knows that the homogenized tensor, which is a deterministic matrix, depends on the solution of a stochastic equation, the so-called corrector problem, which is posed on the whole space $\mathbb{R}^d$. This equation is therefore delicate and expensive to solve. A standard approach consists in truncating the space $\mathbb{R}^d$ to some bounded domain, on which the corrector problem is numerically solved.

In collaboration with B. Stamm (Aachen University, Germany) and S. Xiang (now also at Aachen University, Germany), E. Cancès, V. Ehrlacher and F. Legoll have studied, both from a theoretical and a numerical standpoints, new alternatives for the approximation of the homogenized matrix. They all rely on the use of an embedded corrector problem, previously introduced by the authors, where a finite-size domain made of the highly oscillatory material is embedded in a homogeneous infinite medium whose diffusion coefficients have to be appropriately determined. In [40], they have shown that the different approximations introduced all converge to the homogenized matrix of the medium when the size of the embedded domain goes to infinity. In [41], they present an efficient algorithm for the resolution of such problems for particular heterogeneous materials, based on the reformulation of the embedded corrector problem as an integral equation, which is discretized using spherical harmonics and solved using the fast multipole method.

Besides the averaged behavior of the oscillatory solution $u_\varepsilon$ on large space scales (which is given by its homogenized limit), a question of interest is to describe how $u_\varepsilon$ fluctuates. This question is investigated in the PhD thesis of P.-L. Rothé, both from a theoretical and a numerical viewpoints. First, theoretical results

have been obtained for a weakly stochastic setting (where the coefficient is the sum of a periodic coefficient and a small random perturbation). It has been shown that, at the first order and when $\varepsilon$ is small, the localized fluctuations (characterized by a test function $g$) of $u_\varepsilon$ are Gaussian. The corresponding variance depends on the localization function $g$, on the right-hand side $f$ of the problem satisfied by $u_\varepsilon$, and on a fourth order tensor $Q$ which is defined in terms of the corrector. Since the corrector function is challenging to compute, so is $Q$. A numerical approach has hence been designed to approximate $Q$ and its convergence has been proven. Second, numerical experiments in more general settings (i.e. full stochastic case) following the same approach have been performed. The results are promising, and consistent with the theoretical results obtained in the weakly stochastic setting. These results are collected in a manuscript in preparation.

In collaboration with T. Hudson (University of Warwick, United Kingdom), F. Legoll and T. Lelièvre have considered in [46] a scalar viscoelastic model in which the constitutive law is random and varies on a lengthscale which is small relative to the overall size of the solid. Using stochastic two-scale convergence, they have obtained the homogenized limit of the evolution, and have demonstrated that, under certain hypotheses, the homogenized model exhibits hysteretic behaviour which persists under asymptotically slow loading. This work is motivated by rate-independent stress-strain hysteresis observed in filled rubber.

### *6.3.3. Multiscale Finite Element approaches*

From a numerical perspective, the Multiscale Finite Element Method (MsFEM) is a classical strategy to address the situation when the homogenized problem is not known (e.g. in difficult nonlinear cases), or when the scale of the heterogeneities, although small, is not considered to be zero (and hence the homogenized problem cannot be considered as a sufficiently accurate approximation).

During the year, several research tracks have been pursued in this general direction.

The final writing of the various works performed in the context of the PhD thesis of F. Madiot is still ongoing. The issues examined there are on the one hand the application (and adequate adjustment) of MsFEM approaches to the case of an advection-diffusion equation with a dominating convection term posed in a perforated domain, and on the other hand some more general study of a numerical approach based, again in the case of convection-dominated flows, on the introduction of the invariant measure associated to the problem. The final version of the two manuscripts describing the efforts in each of these directions should be completed in a near future.

The MsFEM approach uses a Galerkin approximation of the problem on a pre-computed basis, obtained by solving local problems mimicking the problem at hand at the scale of mesh elements, with carefully chosen right-hand sides and boundary conditions. The initially proposed version of MsFEM uses as basis functions the solutions to these local problems, posed on each mesh element, with null right-hand sides and with the coarse P1 elements as Dirichlet boundary conditions. Various improvements have next been proposed, such as the *oversampling* variant, which solves local problems on larger domains and restricts their solutions to the considered element. In collaboration with U. Hetmaniuk (University of Washington in Seattle, USA), C. Le Bris, F. Legoll and P.-L. Rothé have introduced and studied a MsFEM method improved differently. They have considered a variant of the classical MsFEM approach with enrichments based on Legendre polynomials, both in the bulk of the mesh elements and on their interfaces. A convergence analysis of this new variant has been performed. Promising numerical results have been obtained. These results are currently being collected in a manuscript in preparation.

One of the perspectives of the team, through the PhD thesis of A. Lesage, is the development of Multiscale Finite Element Methods for thin heterogeneous plates. The fact that one of the dimension of the domain of interest scales as the typical size of the heterogeneities within the material induces theoretical and practical difficulties that have to be carefully taken into account. The first steps of the work of V. Ehrlacher, F. Legoll and A. Lesage, in collaboration with A. Lebée (École des Ponts) have consisted in studying the homogenized limit (and the two-scale expansion) of problems posed on thin heterogeneous plates. The case of a diffusion equation has been first dealt with, while the more challenging case of elasticity is currently under study.

## 6.4. Complex fluids

**Participants:** Sébastien Boyaval, Dena Kazerani.

The aim of the research performed in the project-team about complex fluids is

- to guide the mathematical modeling with PDEs of real materials flows, multi-phase fluids such as suspensions of particles or stratified air-water flows in particular, and
- to propose efficient algorithms for the computation of flow solutions, mainly for the many applications in the hydraulic engineering context.

Concerning the first point, new results have been obtained in collaboration with A. Caboussat (HEG, Switzerland) and M. Picasso (EPFL, Switzerland), in the framework of the SEDIFLO project (funded by ANR) and of Arwa Mrad PhD thesis at EPFL. In [13], they have shown numerically inability of some classical incompressible density-dependent Navier-Stokes equations to take into account some multiphase concentration effects in a prototypical set-up of fluvial erosion (in comparison with physical experiments). Hence the need for *new* models, that better describe complex flows associated with heterogeneities in the fluid microstructure. Concerning the second point, new results have been obtained in collaboration with M. Grepl and K. Veroy (Aachen, Germany) regarding the numerical reduction of transport models for data assimilation [25], in the framework of M. Kaercher PhD thesis at Aachen.

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Contracts and grants with Industry

Many research activities of the project-team are conducted in close collaboration with private or public companies: CEA, SANOFI, EDF. The project-team is also supported by the Office of Naval Research and the European Office of Aerospace Research and Development, for multiscale simulations of random materials. All these contracts are operated at and administrated by the École des Ponts.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

The project-team is involved in several ANR projects:

- S. Boyaval is the PI of the ANR JCJC project SEDIFLO (2016-2020) to investigate new numerical models of solid transport in rivers.
- V. Ehrlacher is a member of the ANR project ADAPT (2018-2022), PI: D. Lombardi, Inria REO team-project. This project is concerned with the parallelization of tensor methods for high-dimensional problems.
- F. Legoll is a member of the ANR project CINE-PARA (2015-2019), PI: Y. Maday, Sorbonne Université. This project is concerned with parallel-in-time algorithms.
- G. Stoltz is the PI of the ANR project COSMOS (2014-2019) which focuses on the development of efficient numerical techniques to simulate high-dimensional systems in molecular dynamics and computational statistics. It includes research teams from Institut Mines-Telecom, Inria Rennes and IBPC Paris.

Members of the project-team are participating in the following GdR:

- AMORE (Advanced Model Order REduction),
- CORREL (correlated methods in electronic structure computations),
- DYNQUA (time evolution of quantum systems, with applications to transport problems, nonequilibrium systems, etc.),
- EGRIN (gravity flows),
- MANU (MAthematics for NUclear applications),
- MASCOT-NUM (stochastic methods for the analysis of numerical codes),
- MEPHY (multiphase flows)
- REST (theoretical spectroscopy),
- CHOCOLAS (experimental and numerical study of shock waves).

The project-team is involved in two Labex: the Labex Bezout (started in 2011) and the Labex MMCD (started in 2012).

## 8.2. European Initiatives

The ERC consolidator Grant MSMATH (ERC Grant Agreement number 614492, PI T. Lelièvre) is running (it started in June 2014).

## 8.3. International Initiatives

T. Lelièvre, G. Stoltz and F. Legoll participate in the Laboratoire International Associé (LIA) CNRS / University of Illinois at Urbana-Champaign on complex biological systems and their simulation by high performance computers. This LIA involves French research teams from Université de Nancy, Institut de Biologie Structurale (Grenoble) and Institut de Biologie Physico-Chimique (Paris). The LIA has been renewed for 4 years, starting January 1st, 2018.

# 9. Dissemination

## 9.1. Promoting Scientific Activities

E. Cancès

- is the director of CERMICS, the Applied Mathematics department at École des Ponts,
- is a member of the editorial boards of Mathematical Modelling and Numerical Analysis (2006-), SIAM Journal of Scientific Computing (2008-), SIAM Multiscale Modeling and Simulation (2012-), and the Journal of Computational Mathematics (2017-),
- has co-organized an Oberwolfach workshop (March), an IMA workshop (March), the 2018 SIAM MS conference (July), and an ISCD summer school (July - August),
- was a member of the DFG Review Panel “Mathematics” for Clusters of Excellence, Cologne, April.

V. Ehrlacher

- is a member of the “Conseil d’Enseignement et de Recherche” of Ecole des Ponts,
- has co-organized the GdR MASCOT-NUM Working meeting on “Uncertainty quantification in materials science”, at IHP, May (with J. Baccou, J. Reygner and G. Perrin).

G. Ferré and J. Roussel have co-organized the working group J-PSI (Jeunes chercheurs en physique statistique et interactions, until July) at IHP. The working group was provided financial support from the SMAI through a BOUM grant, and ended with a one-day conference in June at Inria Paris.

C. Le Bris is a managing editor of Networks and Heterogeneous Media. He is a member of the editorial boards of Annales mathématiques du Québec (2013-), Archive for Rational Mechanics and Analysis (2004-), COCV (Control, Optimization and Calculus of Variations) (2003-), Mathematics in Action (2008-), Nonlinearity (2005-), Journal de Mathématiques Pures et Appliquées (2009-), Pure and Applied Analysis (2018-). He is a member of the editorial boards of the monograph series Mathématiques & Applications, Series, Springer (2008-), Modelling, Simulations and Applications, Series, Springer (2009-), Springer Monographs in Mathematics, Springer (2016-). He is a member of

- the Cabinet of the High Commissioner for Atomic Energy (until September),
- the "International Scientific Advisory Committee" of the Centre de Recherche Mathématique, Université de Montréal (until mid-2018),
- the "Advisory Board" of the DFG Cluster of Excellence Engineering of Advanced Materials, Erlangen,
- the "International Scientific Advisory Board" of the DFG research center Matheon, Berlin,
- the "Conseil scientifique de la SMAI" (Scientific Council of the French Applied Maths Society),
- the International Mathematical Union Circle,
- the "Conseil de la Faculté des sciences et ingénierie", Sorbonne Université.

He is the president of the scientific advisory board of the Institut des Sciences du calcul et des données, Sorbonne Université. He has held a regular position of Visiting Professor at the University of Chicago.

F. Legoll

- is a member of the editorial board of SIAM MMS (2012-) and of ESAIM: Proceedings and Surveys (2012-),
- is a member of the ANR committee CES-40 "mathématiques et informatique".

T. Lelièvre

- is editor-in-chief of ESAIM: Proceedings and Surveys (with D. Chafai, C. Imbert and P. Lafitte),
- is a member of the editorial boards of IMA: Journal of Numerical Analysis and SIAM/ASA Journal of Uncertainty Quantification,
- is a member of the "Conseil d'Administration" of SMAI and École des Ponts,

Together with G. Stoltz, they have

- co-organized the Workshop "Advances in Computational Statistical Physics", CIRM, September (with G. Pavliotis),
- co-organized the CECAM discussion meeting "Coarse-graining with Machine Learning in molecular dynamics", Sanofi Campus Gentilly, December (with P. Gkeka, P. Monmarché).

G. Stoltz

- is a member of the scientific council of UNIT (Université Numérique Ingénierie et Technologie),
- co-organized with C. Robert the workshop "Computational Statistics and Molecular Simulation: A Practical Cross-Fertilization" (BIRS-Oaxaca, November),
- co-organizes the working group "Machine learning and optimization" of the Labex Bezout (with W. Hachel and R. Elie).

### *9.1.1. Conference participation*

Members of the project-team have delivered lectures in the following seminars, workshops and conferences:

- S. Boyaval, weekly seminar of Laboratoire Jean Kuntzmann, Grenoble, February,
- S. Boyaval, GDR EGRIN annual meeting, Clermont-Ferrand, June,
- S. Boyaval, La Trobe University – Kyushu University joint Industrial Math seminar, Melbourne, September,

- E. Cancès, Energy and forces workshop, Cambridge, UK, January,
- E. Cancès, workshop "Mathematical models and computation of nonlinear problems", China, January,
- E. Cancès, weekly seminar of the mathematics department, Sapienza University of Rome, February,
- E. Cancès, weekly seminar of Maison de la Simulation, Saclay, March,
- E. Cancès, 2D materials workshop, Minneapolis, March,
- E. Cancès, Fields Institute workshop, Toronto, May,
- E. Cancès, workshop on computational mathematics, Suzhou, China, June,
- E. Cancès, Centre Henri Lebesgue workshop, Rennes, June,
- E. Cancès, SIAM Materials Science conference, Portland, July,
- E. Cancès, IPAM workshop, Los Angeles, August,
- E. Cancès, GAMM workshop, Aachen, September,
- E. Cancès, Franco-German Meeting Workshop on Mathematical Aspects in Computational Chemistry, Aachen, September,
- E. Cancès, CECAM workshop, Lausanne, November,
- E. Cancès, workshop "Big data challenges for predictive modeling of complex systems", Hong Kong, November,
- V. Ehrlacher, Groupe de travail ENS Rennes, January,
- V. Ehrlacher, workshop on "Mathematical Methods in Quantum Chemistry", Oberwolfach, Germany, March,
- V. Ehrlacher, Séminaire DEFI-MEDISIM-POEMS, October,
- G. Ferré, CERMICS PhD Seminar, Paris, February,
- G. Ferré, Les probabilités de demain, IHP, Paris, March,
- G. Ferré, Congrès National d'Analyse Numérique, Cap d'Agde, May,
- G. Ferré, International Conference in Monte Carlo and Quasi Monte Carlo Methods in Scientific Computing, Rennes, July,
- G. Ferré, SIAM Materials Science conference, Portland, July (two talks),
- G. Ferré, Franco-German Meeting Workshop on Mathematical Aspects in Computational Chemistry, Aachen, September,
- G. Ferré, Student Probability Seminar, Courant Institute of Mathematical Science, New-York, December,
- M. Josien, CANUM conference, Cap d'Agde, May,
- M. Josien, SIAM Materials Science, Portland, USA, July,
- F. Hédin, "PinT 7th Workshop on Parallel-in-Time methods", Roscoff Marine Station, France, May,
- F. Hédin, "CECAM Workshop, Frontiers of coarse graining in molecular dynamics", Zuse Institute Berlin, Germany, July,
- F. Hédin, CIRM Conference "Advances in Computational Statistical Physics", September,
- C. Le Bris, Séminaire Pierre-Louis Lions, Collège de France, January,
- C. Le Bris, Applied Mathematics Colloquium of the University of Maryland, February,
- C. Le Bris, PDE seminar, University of Chicago, April,
- C. Le Bris, Journées de l'Ecole Doctorale Carnot-Pasteur, Université de Besançon, June,
- C. Le Bris, Journées Scientifiques de Marcoule, CEA Marcoule, June,
- C. Le Bris, Journées de Cadarache, CEA Cadarache, June,

- C. Le Bris, (plenary lecture) 25th International Conference on Domain Decomposition Methods, St. John's, Canada, July,
- C. Le Bris, LMS Durham Research Symposium on Homogenization in Disordered Media, Durham, UK, August,
- C. Le Bris, Groupe de travail Calcul des Variations Paris-Ile de France, November
- F. Legoll, EMMC conference, Nantes, March,
- F. Legoll, University of Chicago, CAMP seminar, Chicago, USA, May,
- F. Legoll, AIMS conference, Taipei, Taiwan, July,
- F. Legoll, NumDiff conference, Halle, Germany, September,
- T. Lelièvre, Journée de l'ANR CINE-PARA, Université Paris 13, January,
- T. Lelièvre, Workshop "Interplay of Analysis and Probability in Applied Mathematics", Oberwolfach, Feburary,
- T. Lelièvre, Séminaire de la Maison de la Simulation, Saclay, March,
- T. Lelièvre, Séminaire du LJK, Grenoble, March,
- T. Lelièvre, Séminaire Statistical Machine Learning in Paris, Paris, April,
- T. Lelièvre, Workshop "Data-driven modelling of complex systems", ATI, London, May,
- T. Lelièvre, Workshop "Uncertainty quantification in materials science", IHP, Paris, May,
- T. Lelièvre, Séminaire Mathématiques pour l'Industrie et la Physique, Toulouse, May,
- T. Lelièvre, Fields Institute, "Focus Program on Nanoscale Systems and Coupled Phenomena: Mathematical Analysis, Modeling, and Applications", Toronto, May,
- T. Lelièvre, Workshop "Simulation and probability: recent trends", Rennes, June,
- T. Lelièvre, Workshop "Particle based methods", ICMS, Edinburgh, July,
- T. Lelièvre, CECAM workshop "Frontiers of coarse graining in molecular dynamics", Berlin, July,
- T. Lelièvre, Franco-German Workshop on mathematical aspects in computational chemistry, Aachen, September,
- T. Lelièvre, Séminaire "Simulation, Incertitudes et Méta-modèles", CEA Saclay, October,
- T. Lelièvre, Workshop "Computational Statistics and Molecular Simulation: A Practical Cross-Fertilization", Oaxaca, November,
- T. Lelièvre, Groupe de travail Évolution de Populations et Systèmes de Particules en Interaction, Ecole Polytechnique, December,
- A. Levitt, Mathematical Methods in Quantum Chemistry, Oberwolfach, March,
- A. Levitt, Analytical & Numerical Methods in Quantum Transport, Aalborg, May,
- A. Levitt, Beijing Normal University seminar, June,
- A. Levitt, Chinese Academy of Sciences seminar, June,
- A. Levitt, Franco-German Meeting Workshop on Mathematical Aspects in Computational Chemistry, Aachen September,
- P.-L. Rothé, PhD seminar, Inria Paris, June,
- J. Roussel, SIAM Materials Science conference, Portland, July,
- J. Roussel, Monte Carlo & Quasi-Monte Carlo Methods conference, Rennes, France, July,
- L. Silva Lopes, CECAM Coarse Graining Workshop, Berlin, Germany, July,
- L. Silva Lopes, Advances in Computational Statistical Physics, Marseille, September,
- S. Siraj-Dine, SIAM Materials Science conference, Portland, July,
- G. Stoltz, Seminar of the polymer physics group, ETH Zürich, February,

- G. Stoltz, Applied mathematics seminar Duke University, Durham, North Carolina, USA, February,
- G. Stoltz, Statistical Machine Learning in Paris seminar, Paris, April,
- G. Stoltz, Focus Program on Nanoscale Systems and Coupled Phenomena: Mathematical Analysis, Modeling, and Applications, Fields institute, Toronto, Canada, May,
- G. Stoltz, Journées scientifiques Inria, Bordeaux, France, June,
- G. Stoltz, Applied mathematics seminar Courant Institute of Mathematical Sciences, New York, October,
- G. Stoltz, Inria-LJLL seminar, December,
- P. Terrier, Minerals, Metals & Materials Society Annual Meeting & Exhibition, Phoenix, March,
- P. Terrier, CANUM, Cap d'Adge, June.

Members of the project-team have delivered the following series of lectures:

- E. Cancès, Fourier transform and applications in quantum physics and chemistry, 9h, GDR CORREL spring school, Paris, April,
- E. Cancès, Optimization problems in molecular simulation, 12h, ISCD summer school, Roscoff, July,
- E. Cancès, Mathematical methods and numerical algorithms for quantum chemistry, 12h, MWM autumn school, Gelsenkirchen, October,
- C. Le Bris, Aziz Lectures, University of Maryland, College Park, February,
- C. Le Bris, Fields Institute Coxeter Lecture Series, Toronto, May,
- T. Lelièvre, Mini-school math/chemistry GDR CORREL, 9h, April,
- T. Lelièvre, Lectures on "Stochastic numerical methods and molecular dynamics simulations" (15h), Ecole d'été ISCD (Sorbonne Université), Roscoff, August.

Members of the project-team have presented posters in the following seminars, workshops and international conferences:

- A. Lesage, Fifth workshop on thin structures, Naples, Italy, September,
- J. Roussel, Advances in Computational Statistical Physics, CIRM, Marseille, France, September,
- G. Ferré, Data-driven modelling of Complex Systems, Alan Turing Institute, London,
- G. Ferré, Simulation Aléatoire : problèmes actuels, Inria Rennes,
- G. Ferré, Advances in Computational Statistical physics, CIRM.

Members of the team have benefited from long-term stays in institutions abroad:

- G. Ferré, Courant Institute of Mathematical Science, New York University, New York, USA, October-November,
- P.-L. Rothé, Department of Applied Mathematics, University of Washington, Seattle, USA, April-May.

Members of the project-team have participated (without giving talks nor presenting posters) in the following seminars, workshops and international conferences:

- G. Ferré YES'X Workshop, Scalable Statistics: Accuracy and computational complexity, March,
- M. Josien, Coxeter Lecture Series, Seminar talks, Toronto, Canada, May
- A. Lesage, CANUM conference, Cap d'Agde, May,
- A. Lesage, 6th European conference on computational mechanics, Glasgow, United Kingdom, June,
- M. Ramil, Perspectives en physique statistique computationnelle au CIRM (Centre International de Recherche Mathématiques), Marseille, September,
- M. Ramil, Journées Kolmogorov, Evry, September,
- M. Ramil, ANR EFI workshop, Lyon, November,
- P.-L. Rothé, 6th European conference on computational mechanics, Glasgow, United Kingdom, June,
- P.-L. Rothé, FreeFem++ days, Paris, December,
- S. Siraj-Dine, Workshop Mathematical Challenges in Quantum Mechanics, Rome, February,
- S. Siraj-Dine, Oberwolfach Workshop on Mathematical Methods in Quantum Chemistry, March,
- S. Siraj-Dine, ICMP XIX Congress on Mathematical Physics, Montréal, July.

### *9.1.2. Software development and contributions*

- A. Levitt has implemented a method to construct maximally-localized Wannier functions for metals. A. Levitt and S. Siraj-Dine have implemented a method for the computation of Wannier functions of topological insulators. Both these methods are available at https://github.com/antoine-levitt/wannier.
- J. Roussel and G. Stoltz have restructured the SIMOL code, in particular separating core functions, routines for quantum simulations and advanced features for molecular dynamics, in order to obtain a simpler and more accessible base code. The code is available at https://gitlab.inria.fr/matherials/simol/.
- A first implementation of the Generalized Parallel Replica algorithm, developed by F. Hédin and T. Lelièvre, is available at https://gitlab.inria.fr/parallel-replica/gen.parRep. The objective of the gen.parRep software is to popularize the use of the Parallel Replica algorithm to biological systems. Molecular dynamics is performed by using external codes linked to this program such as OpenMM. This is the first publicly available implementation of the Generalized Parallel Replica method targeting frequently encountered metastable biochemical systems, such as conformational equilibria or dissociation of protein-ligand complexes. We refer to the preprint [22] for more details.

## 9.2. Teaching - Supervision - Juries

The members of the project-team have taught the following courses.

At École des Ponts 1st year (equivalent to L3):

- Analyse et calcul scientifique, 30h (A. Levitt, G. Stoltz),
- Équations aux dérivées partielles et éléments finis, 15h (F. Legoll, P.-L. Rothé),
- Hydraulique numérique, 15h (S. Boyaval),
- Mécanique quantique, 10h (E. Cancès, A. Levitt),
- Méthodes numériques pour les problèmes en grande dimension, 17h30 (V. Ehrlacher, S. Boyaval),
- Optimisation, 15h, L3 (A. Lesage, A. Levitt),
- Outils mathématiques pour l'ingénieur, 15h (E. Cancès, G. Ferré, F. Legoll, T. Lelièvre, P-L. Rothé),
- Probabilités, 27h (M. Ramil)
- Projets de première année, 15h (J. Roussel, P. Terrier),

At École des Ponts 2nd year (equivalent to M1):

- Analyse de Fourier, 15h (A. Levitt),
- Analyse spectrale et application aux Équations aux dérivées partielles, 36h (F. Legoll, V. Ehrlacher),
- Contrôle de systèmes dynamiques et équations aux dérivées partielles, 18h (E. Cancès),
- Projet du département IMI, 12h (G. Ferré, M. Ramil, J. Roussel, L. Silva Lopes),
- Projets Modéliser Programmer Simuler (T. Lelièvre),
- Simulation moléculaire en sciences des matériaux, 6h (A. Levitt),
- Statistics and data sciences, 24h (G. Stoltz).

At École des Ponts 3rd year (equivalent to M2):

- Méthodes de quantification des incertitudes en ingénierie, 18h (V. Ehrlacher),
- Remise à niveau: outils mathématiques, 6h (A. Lesage).

At the M2 "Mathématiques de la modélisation" of Sorbonne Université:

- Introduction à la physique statistique computationnelle, 20h (G. Stoltz),
- Méthodes numériques probabilistes, 24h (T. Lelièvre),
- Problèmes multiéchelles, aspects théoriques et numériques, 24h (F. Legoll),
- Théorie spectrale et variationnelle, 10h (E. Cancès).

At other institutions:

- Analyse variationnelle des équations aux dérivées partielles, 32h, École Polytechnique (T. Lelièvre),
- Aléatoire, 32h, École Polytechnique (T. Lelièvre),
- Maths 1 et 2, 9h, L3, École des Mines (A. Levitt, G. Stoltz),
- Mathématiques pour l'ingénieur, 36h, L2, UPEC (S. Siraj-Dine),
- Numerical methods for partial differential equations, 21h, University of Chicago (C. Le Bris).

The following PhD theses have been defended in the group at École des Ponts:

- Amina Benaceur, Réduction de modèles en thermique et mécanique non-linéaires, Université Paris-Est, École des Ponts, defended on December 21th, 2018, supervised by A. Ern (CERMICS), co-supervised by V. Ehrlacher,
- Marc Josien, Etude mathématique et numérique de quelques modèles multi-échelles issus de la mécanique des matériaux, Université Paris-Est, École des Ponts, defended on November 20th, 2018, supervised by C. Le Bris,
- Julien Roussel, Analyse théorique et numérique de dynamiques non-réversibles en physique statistique computationnelle, Université Paris-Est, École des Ponts, defended on November 27th, 2018, supervised by G. Stoltz,
- Pierre Terrier, Reduced models for defect migration in metals, Université Paris-Est, École des Ponts and CEA Saclay, defended on December 19th, supervised by G. Stoltz and M. Athènes (CEA).

The following PhD theses are ongoing in the group at École des Ponts:

- Zineb Belkacemi, Machine learning techniques in molecular simulation, Université Paris-Est, Thèse CIFRE Sanofi, started November 1st, 2018, co-supervised by T. Lelièvre and G. Stoltz,
- Robert Benda, Multiscale modeling of functionalized nanotube networks for sensor applications, Ecole Polytechnique, started September 1st, 2018, supervised by E. Cancès and B. Lebental (École Polytechnique),
- Raed Blel, Monte Carlo methods and model redcution, started October 1st, 2018, supervised by V. Ehrlacher and T. Lelièvre,
- Lingling Cao, Mathematical analysis of models of thermo-electronic transport, Université Paris-Est, École des Ponts, started November 1st, 2016, supervised by E. Cancès and G. Stoltz,
- Rafaël Coyaud, Méthodes numériques déterministes et stochastiques pour le transport optimal, Université Paris-Est, École des Ponts, started October 1st, 2017, supervised by A. Alfonsi (CERMICS) and co-supervised by V. Ehrlacher,
- Qiming Du, Mathematical analysis of splitting methods, École Doctorale Sciences Mathématiques de Paris Centre, started September 1st, 2016, supervised by A. Guyader (Sorbonne Université) and T. Lelièvre,
- Grégoire Ferré, Efficient sampling methods for nonequilibrium systems, Université Paris-Est, École des Ponts started October 1st, 2016, supervised by G. Stoltz,
- Adrien Lesage, Multi-scale methods for calculation and optimization of thin structures, started October 1st, 2017, supervised by F. Legoll, co-supervised by V. Ehrlacher and A. Lebée (École des Ponts),
- Sofiane Martel, Modélisation de la turbulence par mesures invariantes d'EDPS, Université Paris-Est, École des Ponts, started January 1st, 2017, supervised by S. Boyaval and co-supervised by J. Reygner (CERMICS),
- Pierre-Loïk Rothé, Numerical methods for the estimation of fluctuations in multi-scale materials and related problems, started October 1st, 2016, supervised by F. Legoll,
- Mouad Ramil, Metastability for interacting particle systems, started October 1st, 2017, supervised by T. Lelièvre and J. Reygner (CERMICS),
- Laura Silva Lopes, Numerical methods for simulating rare events in molecular dynamics, started October 1st, 2016, supervised by J. Hénin (IBPC) and T. Lelièvre,
- Sami Siraj-Dine, Modélisation mathématique des matériaux 2D, École des Ponts, started October 1st, 2017, supervised by E. Cancès, C. Fermanian and co-supervised by A. Levitt.

Project-team members have participated in the following PhD juries:

- E. Cancès, PhD of Marco Vanzini ("Auxiliary systems for observables: dynamical local connector approximation for electron addition and removal spectra"), defended at Ecole Polytechnique in January 2018,
- E. Cancès, PhD of Giovanna Marcelli ("A mathematical analysis of spin and charge transport in topological insulators"), defended at Sapienza University of Rome in February 2018,
- E. Cancès, PhD of Mi-Song Dupuy ("Analyse de la méthode projector augmented-wave pour les calculs de structure électronique en géométrie périodique"), defended at Université Paris Diderot in September 2018,
- E. Cancès, PhD of Carlo Marcati ("Discontinuous hp finite element methods for elliptic eigenvalue problems with singular potentials, with applications in quantum chemistry"), defended at Sorbonne Université in October 2018,
- V. Ehrlacher, PhD of Mi-Song Dupuy, ("Analyse de la méthode projector augmented-wave pour les calculs de structure électronique en géométrie périodique"), defended at Université Paris-Diderot in September 2018.
- V. Ehrlacher, PhD of Nicolas Cagniart, ("Quelques approches non linéaires en réduction de complexité"), defended at Sorbonne Université in November 2018,
- V. Ehrlacher, PhD of Jules Fauque, ("Modèle d'ordre réduit en mécanique du contact. Application à la simulation du comportement des combustibles nucléaires"), defended at Ecole des Mines de Paris in November 2018,
- V. Ehrlacher, PhD of Ahmad Al-Takash, ("Development of numerical methods to accelerate the prediction of the behavior of multiphysics under cyclic loading"), defended at ENSMA in November 2018,
- F. Legoll, PhD of Brian Staber ("Stochastic analysis, simulation and identification of hyperelastic constitutive equations"), defended at Université Paris-Est in June 2018,
- T. Lelièvre, PhD of Bob Pépin ("Time Averages of Diffusion Processes and Applications to Two-Timescale Problems"), défended at Université du Luxembourg, April 2018,
- T. Lelièvre, PhD of Michel Nowak ("Accelerating Monte Carlo particle transport with adaptively generated importance maps"), defended at Université Paris Saclay, September 2018,
- T. Lelièvre, PhD of Ze Lei ("Irreversible Markov Chains for Particle Systems and Spin Models: Mixing and Dynamical Scaling"), défended at Ecole Normale Supérieure, December 2018,
- G. Stoltz, PhD of Sabri Souguir ("Simulation numérique de l'initiation de la rupture à l'échelle atomique"), defended at Ecole des Ponts in November 2018.

## 9.3. Popularization

### *9.3.1. Internal or external Inria responsibilities*

- A. Levitt is a member of the editorial board of Interstices, Inria's popularization website.

### *9.3.2. Articles and contents*

- E. Cancès has been interviewed in "La Jaune et La Rouge", the journal of the alumni of Ecole Polytechnique, in January.

### *9.3.3. Internal actions*

- C. Le Bris organized an open day at CERMICS in June for the administrative staff of École des Ponts.

# 10. Bibliography

## Major publications by the team in recent years

[1] E. CANCÈS, M. DEFRANCESCHI, W. KUTZELNIGG, C. LE BRIS, Y. MADAY.*Computational Quantum Chemistry: A Primer*, 2003, Le Bris, Claude (ed.), Special Volume: Computational Chemistry. Amsterdam: North-Holland. Handb. Numer. Anal. 10, 3-270 (2003)

[2] E. CANCÈS, C. LE BRIS, Y. MADAY.*Mathematical Methods in Quantum Chemistry. An Introduction. (Méthodes mathématiques en chimie quantique. Une introduction.)*, Mathématiques et Applications (Berlin) 53. Berlin: Springer. xvi, 409 p. , 2006

[3] I. CATTO, C. LE BRIS, P.-L. LIONS.*The Mathematical Theory of Thermodynamic Limits: Thomas-Fermi Type Models*, Oxford Mathematical Monographs. Oxford: Clarendon Press. xiii, 277 p., 1998

[4] J.-F. GERBEAU, C. LE BRIS, T. LELIÈVRE.*Mathematical Methods for the Magnetohydrodynamics of Liquid Metals*, Numerical Mathematics and Scientific Computation. Oxford: Oxford University Press., 324 p., 2006

[5] C. LE BRIS.*Multi-scale Analysis. Modeling and Simulation. (Systèmes multi-échelles. Modélisation et simulation.)*, Mathématiques et Applications (Berlin) 47. Berlin: Springer. xi, 212 p., 2005

[6] T. LELIÈVRE, M. ROUSSET, G. STOLTZ.*Free Energy Computations: A Mathematical Perspective*, Imperial College Press, 458 p., 2010

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[7] M. JOSIEN.*Mathematical and numerical study of some multi-scale models from materials science*, Université Paris-Est, November 2018, https://hal.archives-ouvertes.fr/tel-01988719

[8] J. ROUSSEL.*Theoretical and numerical analysis of non-reversible dynamics in computational statistical physics+*, MSTIC graduate school / University of Marne-la-vallée, November 2018, https://tel.archives-ouvertes.fr/tel-01964722

[9] P. TERRIER.*Numerical simulations for predicting the microstructural evolution of ferritic alloys. A study of Cluster Dynamics*, Université Paris-Est, December 2018, https://tel.archives-ouvertes.fr/tel-01990556

### Articles in International Peer-Reviewed Journal

[10] A. BENACEUR, V. EHRLACHER, A. ERN, S. MEUNIER.*A progressive reduced basis/empirical interpolation method for nonlinear parabolic problems*, in "SIAM Journal on Scientific Computing", 2018, vol. 40, n$^{o}$ 5, p. A2930-A2955, https://arxiv.org/abs/1710.00511 , https://hal.archives-ouvertes.fr/hal-01599304

[11] X. BLANC, M. JOSIEN, C. LE BRIS.*Local precised approximation in multiscale problems with local defects*, in "Comptes Rendus Mathématique", 2019, https://arxiv.org/abs/1901.09669 [*DOI :* 10.1016/J.CRMA.2018.12.005], https://hal.archives-ouvertes.fr/hal-01893991

[12] X. BLANC, P.-L. LIONS, C. LE BRIS.*On correctors for linear elliptic homogenization in the presence of local defects: the case of advection-diffusion*, in "Journal de Mathématiques Pures et Appliquées", 2018, https://arxiv.org/abs/1801.10330 [*DOI :* 10.1016/J.MATPUR.2018.04.010], https://hal.archives-ouvertes.fr/hal-01697105

[13] S. BOYAVAL, A. CABOUSSAT, A. MRAD, M. PICASSO, G. STEINER.*A semi-Lagrangian splitting method for the numerical simulation of sediment transport with free surface flows*, in "Computers & Fluids", August 2018, vol. 172, p. 384-396, https://hal-enpc.archives-ouvertes.fr/hal-01940147

[14] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Guaranteed and robust a posteriori bounds for Laplace eigenvalues and eigenvectors: a unified framework*, in "Numerische Mathematik", July 2018, vol. 140, n$^{o}$ 4, p. 1033-1079 [*DOI :* 10.1007/S00211-018-0984-0], https://hal.inria.fr/hal-01483461

[15] P. CARDALIAGUET, C. LE BRIS, P. E. SOUGANIDIS.*Perturbation problems in homogenization of hamilton-jacobi equations*, in "Journal de Mathématiques Pures et Appliquées", 2019, https://arxiv.org/abs/1701.05440 , https://hal.archives-ouvertes.fr/hal-01435744

[16] D. CHAFAÏ, G. FERRÉ.*Simulating Coulomb gases and log-gases with hybrid Monte Carlo algorithms*, in "Journal of Statistical Physics", November 2018, https://arxiv.org/abs/1806.05985 [*DOI :* 10.1007/S10955-018-2195-6], https://hal.archives-ouvertes.fr/hal-01818268

[17] L. CHAMOIN, F. LEGOLL.*A posteriori error estimation and adaptive strategy for the control of MsFEM computations*, in "Computer Methods in Applied Mechanics and Engineering", 2018, vol. 336, p. 1-38, https://arxiv.org/abs/1709.03624 [*DOI :* 10.1016/J.CMA.2018.02.016], https://hal.archives-ouvertes.fr/hal-01586825

[18] G. FAURE, G. STOLTZ.*Stable and accurate schemes for smoothed dissipative particle dynamics*, in "Applied Mathematics and Mechanics", January 2018, vol. 39, n$^{o}$ 1, p. 83-102, https://arxiv.org/abs/1707.04232 [*DOI :* 10.1007/S10483-018-2256-8], https://hal.archives-ouvertes.fr/hal-01562490

[19] G. FERRÉ, H. TOUCHETTE.*Adaptive Sampling of Large Deviations*, in "Journal of Statistical Physics", September 2018, vol. 172, n$^{o}$ 6, p. 1525 - 1544 [*DOI :* 10.1007/S10955-018-2108-8], https://hal.archives-ouvertes.fr/hal-01938825

[20] G. FORT, B. JOURDAIN, T. LELIÈVRE, G. STOLTZ.*Convergence and efficiency of adaptive importance sampling techniques with partial biasing*, in "Journal of Statistical Physics", March 2018, vol. 171, n$^{o}$ 2, p. 220–268, https://arxiv.org/abs/1610.09194 [*DOI :* 10.1007/S10955-018-1992-2], https://hal.archives-ouvertes.fr/hal-01389996

[21] C. L. HALL, T. HUDSON, P. V. MEURS.*Asymptotic analysis of boundary layers in a repulsive particle system*, in "Acta Applicandae Mathematicae", February 2018, https://arxiv.org/abs/1609.03236 [*DOI :* 10.1007/S10440-017-0119-0], https://hal.archives-ouvertes.fr/hal-01398449

[22] F. HÉDIN, T. LELIÈVRE.*gen.parRep: a first implementation of the Generalized Parallel Replica dynamics for the long time simulation of metastable biochemical systems*, in "Computer Physics Communications", 2019, https://arxiv.org/abs/1807.02431 - Accepte version of the article. 37 pages (including references), 13 Figures, 3 Tables [*DOI :* 10.1016/J.CPC.2019.01.005], https://hal.archives-ouvertes.fr/hal-01832823

[23] J. INFANTE ACEVEDO, T. LELIÈVRE.*A non linear approximation method for solving high dimensional partial differential equations: Application in finance*, in "Mathematics and Computers in Simulation", January 2018, vol. 143, p. 14-34, https://arxiv.org/abs/1309.3731 [*DOI :* 10.1016/J.MATCOM.2016.07.013], https://hal-enpc.archives-ouvertes.fr/hal-00861892

[24] M. JOSIEN, Y.-P. PELLEGRINI, F. LEGOLL, C. LE BRIS.*Fourier-based numerical approximation of the Weertman equation for moving dislocations*, in "International Journal for Numerical Methods in Engineering", 2018, vol. 113, n$^{o}$ 12, p. 1827-1850, https://arxiv.org/abs/1704.04489 [*DOI :* 10.1002/NME.5723], https://hal.archives-ouvertes.fr/hal-01510158

[25] M. KAERCHER, S. BOYAVAL, M. GREPL, K. VEROY.*Reduced basis approximation and a posteriori error bounds for 4D-Var data assimilation*, in "Optimization and Engineering", June 2018, vol. 2018, n$^{o}$ 3, https://arxiv.org/abs/1802.02328 [*DOI :* 10.1007/S11081-018-9389-2], https://hal.inria.fr/hal-01556304

[26] C. LE BRIS, F. LEGOLL, S. LEMAIRE.*On the best constant matrix approximating an oscillatory matrix-valued coefficient in divergence-form operators*, in "ESAIM: Control, Optimisation and Calculus of Variations", 2018, https://arxiv.org/abs/1612.05807 [*DOI :* 10.1051/COCV/2017061], https://hal.archives-ouvertes.fr/hal-01420187

[27] T. LELIÈVRE, L. PILLAUD-VIVIEN, J. REYGNER.*Central Limit Theorem for stationary Fleming–Viot particle systems in finite spaces*, in "ALEA : Latin American Journal of Probability and Mathematical Statistics", September 2018, vol. 15, p. 1163-1182, https://arxiv.org/abs/1806.04490 [*DOI :* 10.30757/ALEA.V15-43], https://hal-enpc.archives-ouvertes.fr/hal-01812120

[28] J. ROUSSEL, G. STOLTZ.*Spectral methods for Langevin dynamics and associated error estimates*, in "ESAIM: Mathematical Modelling and Numerical Analysis", September 2018, vol. 52, n$^{o}$ 3, p. 1051-1083, https://arxiv.org/abs/1702.04718 [*DOI :* 10.1051/M2AN/2017044], https://hal.archives-ouvertes.fr/hal-01470251

[29] B. STAMM, L. LAGARDÈRE, G. SCALMANI, P. GATTO, E. CANCÈS, J.-P. PIQUEMAL, Y. MADAY, B. MENNUCCI, F. LIPPARINI.*How to make continuum solvation incredibly fast in a few simple steps: a practical guide to the domain decomposition paradigm for the Conductor-like Screening Model Continuum Solvation, Linear Scaling, Domain Decomposition*, in "International Journal of Quantum Chemistry", 2018 [*DOI :* 10.1002/QUA.25669], https://hal.archives-ouvertes.fr/hal-01793770

[30] G. STOLTZ, Z. TRSTANOVA.*Langevin dynamics with general kinetic energies*, in "Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal", May 2018, vol. 16, n$^{o}$ 2, p. 777-806, https://arxiv.org/abs/1609.02891 [*DOI :* 10.1137/16M110575X], https://hal.archives-ouvertes.fr/hal-01364821

[31] G. STOLTZ, E. VANDEN-EIJNDEN.*Longtime convergence of the Temperature-Accelerated Molecular Dynamics Method*, in "Nonlinearity", July 2018, vol. 31, n$^{o}$ 8, p. 3748-3769, https://arxiv.org/abs/1708.08800 [*DOI :* 10.1088/1361-6544/AAC541], https://hal.archives-ouvertes.fr/hal-01578911

## Other Publications

[32] C. ANDRIEU, A. DURMUS, N. NÜSKEN, J. ROUSSEL.*Hypocoercivity of Piecewise Deterministic Markov Process-Monte Carlo*, December 2018, https://arxiv.org/abs/1808.08592 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01941467

[33] X. BLANC, M. JOSIEN, C. LE BRIS.*Precised approximations in elliptic homogenization beyond the periodic setting*, December 2018, https://arxiv.org/abs/1812.07220 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01958207

[34] X. BLANC, C. LE BRIS, P.-L. LIONS.*On correctors for linear elliptic homogenization in the presence of local defects*, January 2018, https://arxiv.org/abs/1801.10335 - working paper or preprint [*DOI :* 10.1080/03605302.2018.1484764], https://hal.archives-ouvertes.fr/hal-01697104

[35] T. BOIVEAU, V. EHRLACHER, A. ERN, A. NOUY.*Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods*, October 2018, https://arxiv.org/abs/1712.07256 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01668316

[36] C.-E. BRÉHIER, T. LELIÈVRE.*On a new class of score functions to estimate tail probabilities of some stochastic processes with Adaptive Multilevel Splitting*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01923385

[37] E. CANCÈS, L.-L. CAO, G. STOLTZ.*Removing a slab from the Fermi sea: the reduced Hartree-Fock model*, October 2018, https://arxiv.org/abs/1807.06960 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01891488

[38] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Post-processing of the planewave approximation of Schrödinger equations. Part I: linear operators*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01908039

[39] E. CANCÈS, V. EHRLACHER, D. GONTIER, A. LEVITT, D. LOMBARDI.*Numerical quadrature in the Brillouin zone for periodic Schrödinger operators*, May 2018, https://arxiv.org/abs/1805.07144 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01796582

[40] E. CANCÈS, V. EHRLACHER, F. LEGOLL, B. STAMM, S. XIANG.*An embedded corrector problem for homogenization. Part I: Theory*, July 2018, https://arxiv.org/abs/1807.05131 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01840993

[41] E. CANCÈS, V. EHRLACHER, F. LEGOLL, B. STAMM, S. XIANG.*An embedded corrector problem for homogenization. Part II: Algorithms and discretization*, October 2018, https://arxiv.org/abs/1810.09885 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01903486

[42] A. DAMLE, A. LEVITT, L. LIN.*Variational formulation for Wannier functions with entangled band structure*, January 2018, https://arxiv.org/abs/1801.08572 - 26 pages, 10 figures, https://hal.archives-ouvertes.fr/hal-01696529

[43] G. FERRÉ, M. ROUSSET, G. STOLTZ.*More on the long time stability of Feynman-Kac semigroups*, November 2018, https://arxiv.org/abs/1807.00390 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01911620

[44] G. FERRÉ, G. STOLTZ.*Error estimates on ergodic properties of discretized Feynman-Kac semigroups*, January 2018, https://arxiv.org/abs/1712.04013 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01690532

[45] D. GONTIER, A. LEVITT, S. SIRAJ-DINE.*Numerical construction of Wannier functions through homotopy*, December 2018, https://arxiv.org/abs/1812.06746 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01955842

[46] T. J. HUDSON, F. LEGOLL, T. LELIÈVRE.*Stochastic homogenization of a scalar viscoelastic model exhibiting stress-strain hysteresis*, February 2018, https://arxiv.org/abs/1802.05549 - 35 pages, 4 figures, https://hal.archives-ouvertes.fr/hal-01710772

[47] M. JOSIEN.*Decomposition and pointwise estimates of periodic Green functions of some elliptic equations with periodic oscillatory coefficients*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01848268

[48] F. LEGOLL, T. LELIÈVRE, U. SHARMA.*Effective dynamics for non-reversible stochastic differential equations: a quantitative study*, September 2018, https://arxiv.org/abs/1809.10498 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01883330

[49] T. LELIÈVRE, D. LE PEUTREC, B. NECTOUX.*Exit event from a metastable state and Eyring-Kramers law for the overdamped Langevin dynamics*, November 2018, https://arxiv.org/abs/1811.06786 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01696187

[50] T. LELIÈVRE.*Mathematical foundations of Accelerated Molecular Dynamics methods*, January 2018, https://arxiv.org/abs/1801.05347 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01686062

[51] T. LELIÈVRE, M. ROUSSET, G. STOLTZ.*Hybrid Monte Carlo methods for sampling probability measures on submanifolds*, July 2018, https://arxiv.org/abs/1807.02356 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01832820

[52] T. LELIÈVRE, G. SAMAEY, P. ZIELIŃSKI.*Analysis of a micro-macro acceleration method with minimum relative entropy moment matching*, January 2018, https://arxiv.org/abs/1801.01740 - 40 pages, https://hal.archives-ouvertes.fr/hal-01680761

[53] T. LELIÈVRE, W. ZHANG.*Pathwise estimates for effective dynamics: the case of nonlinear vectorial reaction coordinates*, May 2018, https://arxiv.org/abs/1805.01928 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01794919

[54] A. LEVITT.*Screening in the finite-temperature reduced Hartree-Fock model*, October 2018, https://arxiv.org/abs/1810.03342 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01889411

[55] J. ROUSSEL, G. STOLTZ.*A perturbative approach to control variates in molecular dynamics*, February 2018, https://arxiv.org/abs/1712.08022 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01709265

# Project-Team MATHRISK

## Mathematical Risk handling

IN COLLABORATION WITH: Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique (CERMICS)

IN PARTNERSHIP WITH:
**CNRS**

**Ecole des Ponts ParisTech**

**Université Paris-Est Marne-la-Vallée**

RESEARCH CENTER
**Paris**

THEME
**Stochastic approaches**

# Table of contents

## Project-Team MATHRISK

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

**Keywords:**

**Computer Science and Digital Science:**

A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.2. - Stochastic Modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.2. - Numerical probability
A6.2.3. - Probabilistic methods
A6.4.2. - Stochastic control

**Other Research Topics and Application Domains:**

B3.1. - Sustainable development
B9.6.3. - Economy, Finance
B9.11. - Risk management

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Agnès Bialobroda Sulem [Team leader, Inria, Senior Researcher, HDR]
Aurélien Alfonsi [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Bernard Lapeyre [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]
Benjamin Jourdain [Ecole Nationale des Ponts et Chaussées, Senior Researcher, HDR]

**Faculty Members**

Vlad Bally [Univ Paris-Est Marne La Vallée, Professor, HDR]
Damien Lamberton [Univ Paris-Est Marne La Vallée, Professor, HDR]

**External Collaborators**

Oleg Kudryavtsev [Roskov Univ, from Jun 2018 until Aug 2018]
Céline Labart [Univ de Savoie]
Jérôme Lelong [ENSIMAG, HDR]
Antonino Zanette [University of Udine, HDR]

**Technical Staff**

Cédric Doucet [Inria, until Mar 2018]
Pierre-Guillaume Raverdy [Inria, from Apr 2018]

**PhD Students**

Oumaima Bencheikh [Ecole Nationale des Ponts et Chaussées]
Adel Cherchali [Ecole Nationale des Ponts et Chaussées]
Alexandre Zhou [Ecole Nationale des Ponts et Chaussées]
Giulia Terenzi [Univ Paris-Est Marne La Vallée]
Marouan Iben Taarit [Ecole Nationale des Ponts et Chaussées]
Rafael Coyaud [Ecole Polytechnique]
William Margheriti [Ecole Nationale des Ponts et Chaussées]
Ezechiel Kahn [Ecole Nationale des Ponts et Chaussées, from Sep 2018]

**Administrative Assistants**

Derya Gok [Inria, from Sep 2018]
Martine Verneuille [Inria, until Sep 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

The Inria project team **MathRisk** team was created in 2013. It is the follow-up of the MathFi project team founded in 2000. MathFi was focused on financial mathematics, in particular on computational methods for pricing and hedging increasingly complex financial products. The 2007 global financial crisis and its "aftermath crisis" has abruptly highlighted the critical importance of a better understanding and management of risk. The project **MathRisk** has been reoriented towards mathematical handling of risk, and addresses broad research topics embracing risk measurement and risk management, modeling and optimization in quantitative finance, but also in other related domains where risk control is paramount. The project team **MathRisk** aims both at producing mathematical tools and models in these domains, and developing collaborations with various institutions involved in risk control. Quantitative finance remains for the project an important source of mathematical problems and applications. Indeed, the pressure of new legislation leads to a massive reorientation of research priorities, and the interest of analysts shifted to risk control preoccupation.

The scientific issues related to quantitative finance we consider include systemic risk and contagion modeling, robust finance, market frictions, counterparty and liquidity risk, assets dependence modeling, market microstructure modeling and price impact. In this context, models must take into account the multidimensional feature and various market imperfections. They are much more demanding mathematically and numerically, and require the development of risk measures taking into account incompleteness issues, model uncertainties, interplay between information and performance and various defaults.

Besides, financial institutions, submitted to more stringent regulatory legislations such as FRTB or XVA computation, are facing practical implementation challenges which still need to be solved. Research focused on numerical efficiency remains strongly needed in this context, renewing the interest for the numerical platform Premia (http://www.premia.fr) that Mathrisk is developing in collaboration with a consortium of financial institutions.

While these themes arise naturally in the world of quantitative finance, a number of these issues and mathematical tools are also relevant to the treatment of risk in other areas as economy, social insurance and sustainable development, of fundamental importance in today's society. In these contexts, the management of risk appears at different time scales, from high frequency data to long term life insurance management, raising challenging renewed modeling and numerical issues.

The **MathRisk** project is strongly involved in the development of new mathematical methods and numerical algorithms. Mathematical tools include stochastic modeling, stochastic analysis, in particular stochastic (partial) differential equations and various aspects of stochastic control and optimal stopping of these equations, nonlinear expectations, Malliavin calculus, stochastic optimization, dynamic game theory, random graphs, martingale optimal transport (especially in relation to numerical considerations), long time behavior of Markov processes (with applications to Monte-Carlo methods) and generally advanced numerical methods for effective solutions.

# 3. Research Program

## 3.1. Risk management: modeling and optimization

### *3.1.1. Contagion modeling and systemic risk*

After the recent financial crisis, systemic risk has emerged as one of the major research topics in mathematical finance. Interconnected systems are subject to contagion in time of distress. The scope is to understand and model how the bankruptcy of a bank (or a large company) may or not induce other bankruptcies. By contrast with the traditional approach in risk management, the focus is no longer on modeling the risks faced by a single financial institution, but on modeling the complex interrelations between financial institutions and the mechanisms of distress propagation among these.

The mathematical modeling of default contagion, by which an economic shock causing initial losses and default of a few institutions is amplified due to complex linkages, leading to large scale defaults, can be addressed by various techniques, such as network approaches (see in particular R. Cont et al. [40] and A. Minca [79]) or mean field interaction models (Garnier-Papanicolaou-Yang [70]).

We have contributed in the last years to the research on the control of contagion in financial systems in the framework of random graph models : In [41], [80], [5], A. Sulem with A. Minca and H. Amini consider a financial network described as a weighted directed graph, in which nodes represent financial institutions and edges the exposures between them. The distress propagation is modeled as an epidemics on this graph. They study the optimal intervention of a lender of last resort who seeks to make equity infusions in a banking system prone to insolvency and to bank runs, under complete and incomplete information of the failure cluster, in order to minimize the contagion effects. The paper [5] provides in particular important insight on the relation between the value of a financial system, connectivity and optimal intervention.

The results show that up to a certain connectivity, the value of the financial system increases with connectivity. However, this is no longer the case if connectivity becomes too large. The natural question remains how to create incentives for the banks to attain an optimal level of connectivity. This is studied in [54], where network formation for a large set of financial institutions represented as nodes is investigated. Linkages are source of income, and at the same time they bear the risk of contagion, which is endogeneous and depends on the strategies of all nodes in the system. The optimal connectivity of the nodes results from a game. Existence of an equilibrium in the system and stability properties is studied. The results suggest that financial stability is best described in terms of the mechanism of network formation than in terms of simple statistics of the network topology like the average connectivity.

### *3.1.2. Liquidity risk and Market Microstructure*

Liquidity risk is the risk arising from the difficulty of selling (or buying) an asset. Usually, assets are quoted on a market with a Limit Order Book (LOB) that registers all the waiting limit buy and sell orders for this asset. The bid (resp. ask) price is the most expensive (resp. cheapest) waiting buy or sell order. If a trader wants to sell a single asset, he will sell it at the bid price, but if he wants to sell a large quantity of assets, he will have to sell them at a lower price in order to match further waiting buy orders. This creates an extra cost, and raises important issues. From a short-term perspective (from few minutes to some days), it may be interesting to split the selling order and to focus on finding optimal selling strategies. This requires to model the market microstructure, i.e. how the market reacts in a short time-scale to execution orders. From a long-term perspective (typically, one month or more), one has to understand how this cost modifies portfolio managing strategies (especially delta-hedging or optimal investment strategies). At this time-scale, there is no need to model precisely the market microstructure, but one has to specify how the liquidity costs aggregate.

For rather liquid assets, liquidity risk is usually taken into account via price impact models which describe how a (large) trader influences the asset prices. Then, one is typically interested in the optimal execution problem: how to buy/sell a given amount of assets optimally within a given deadline. This issue is directly related to the existence of statistical arbitrage or Price Manipulation Strategies (PMS). Most of price impact models deal with single assets. A. Alfonsi, F. Klöck and A. Schied [39] have proposed a multi-assets price impact model that extends previous works. Price impact models are usually relevant when trading at an intermediary frequency (say every hour). At a lower frequency, price impact is usually ignored while at a high frequency (every minute or second), one has to take into account the other traders and the price jumps, tick by tick. Midpoint price models are thus usually preferred at this time scale. With P. Blanc, Alfonsi [3] has proposed a model that makes a bridge between these two types of model: they have considered an (Obizhaeva and

Wang) price impact model, in which the flow of market orders generated by the other traders is given by an exogeneous process. They have shown that Price Manipulation Strategies exist when the flow of order is a compound Poisson process. However, modeling this flow by a mutually exciting Hawkes process with a particular parametrization allows them to exclude these PMS. Besides, the optimal execution strategy is explicit in this model. A practical implementation is given in [35].

### 3.1.3. Dependence modeling

**- Calibration of stochastic and local volatility models.** The volatility is a key concept in modern mathematical finance, and an indicator of market stability. Risk management and associated instruments depend strongly on the volatility, and volatility modeling is a crucial issue in the finance industry. Of particular importance is the assets *dependence* modeling.

By Gyongy's theorem, a local and stochastic volatility model is calibrated to the market prices of all call options with positive maturities and strikes if its local volatility function is equal to the ratio of the Dupire local volatility function over the root conditional mean square of the stochastic volatility factor given the spot value. This leads to a SDE nonlinear in the sense of McKean. Particle methods based on a kernel approximation of the conditional expectation, as presented by Guyon and Henry-Labordère [71], provide an efficient calibration procedure even if some calibration errors may appear when the range of the stochastic volatility factor is very large. But so far, no existence result is available for the SDE nonlinear in the sense of McKean. In the particular case when the local volatility function is equal to the inverse of the root conditional mean square of the stochastic volatility factor multiplied by the spot value given this value and the interest rate is zero, the solution to the SDE is a fake Brownian motion. When the stochastic volatility factor is a constant (over time) random variable taking finitely many values and the range of its square is not too large, B. Jourdain and A. Zhou proved existence to the associated Fokker-Planck equation [77]. Thanks to results obtained by Figalli in [63], they deduced existence of a new class of fake Brownian motions. They extended these results to the special case of the LSV model called Regime Switching Local Volatility, when the stochastic volatility factor is a jump process taking finitely many values and with jump intensities depending on the spot level.

**- Interest rates modeling.** Affine term structure models have been popularized by Dai and Singleton [55], Duffie, Filipovic and Schachermayer [56]. They consider vector affine diffusions (the coordinates are usually called factors) and assume that the short interest rate is a linear combination of these factors. A model of this kind is the Linear Gaussian Model (LGM) that considers a vector Ornstein-Uhlenbeck diffusions for the factors, see El Karoui and Lacoste [62]. A. Alfonsi et al. [33] have proposed an extension of this model, when the instantaneous covariation between the factors is given by a Wishart process. Doing so, the model keeps its affine structure and tractability while generating smiles for option prices. A price expansion around the LGM is obtained for Caplet and Swaption prices.

### 3.1.4. Robust finance

**- Numerical Methods for Martingale Optimal Transport problems.**

The Martingale Optimal Transport (MOT) problem introduced in [53] has received a recent attention in finance since it gives model-free hedges and bounds on the prices of exotic options. The market prices of liquid call and put options give the marginal distributions of the underlying asset at each traded maturity. Under the simplifying assumption that the risk-free rate is zero, these probability measures are in increasing convex order, since by Strassen's theorem this property is equivalent to the existence of a martingale measure with the right marginal distributions. For an exotic payoff function of the values of the underlying on the time-grid given by these maturities, the model-free upper-bound (resp. lower-bound) for the price consistent with these marginal distributions is given by the following martingale optimal transport problem : maximize (resp. minimize) the integral of the payoff with respect to the martingale measure over all martingale measures with the right marginal distributions. Super-hedging (resp. sub-hedging) strategies are obtained by solving the dual problem. With J. Corbetta, A. Alfonsi and B. Jourdain [36] have studied sampling methods preserving the convex order for two probability measures $\mu$ and $\nu$ on $\mathbf{R}^d$, with $\nu$ dominating $\mu$.

Their method is the first generic approach to tackle the martingale optimal transport problem numerically and can also be applied to several marginals.

**- Robust option pricing in financial markets with imperfections.**

A. Sulem, M.C. Quenez and R. Dumitrescu have studied robust pricing in an imperfect financial market with default. The market imperfections are taken into account via the nonlinearity of the wealth dynamics. In this setting, the pricing system is expressed as a nonlinear g-expectation $\mathcal{E}^g$ induced by a nonlinear BSDE with nonlinear driver $g$ and default jump (see [24]). A large class of imperfect market models can fit in this framework, including imperfections coming from different borrowing and lending interest rates, taxes on profits from risky investments, or from the trading impact of a large investor seller on the market prices and the default probability. Pricing and superhedging issues for American and game options in this context and their links with optimal stopping problems and Dynkin games with nonlinear expectation have been studied. These issues have also been addressed in the case of model uncertainty, in particular uncertainty on the default probability. The seller's robust price of a game option has been characterized as the value function of a Dynkin game under $\mathcal{E}^g$ expectation as well as the solution of a nonlinear doubly reflected BSDE in [9]. Existence of robust superhedging strategies has been studied. The buyer's point of view and arbitrage issues have also been studied in this context.

In a Markovian framework, the results of the paper [8] on combined optimal stopping/stochastic control with $\mathcal{E}^g$ expectation allows us to address American nonlinear option pricing when the payoff function is only Borelian and when there is ambiguity both on the drift and the volatility of the underlying asset price process. Robust optimal stopping of dynamic risk measures induced by BSDEs with jumps with model ambiguity is studied in [82].

## 3.2. Perspectives in Stochastic Analysis

### *3.2.1. Optimal transport and longtime behavior of Markov processes*

The dissipation of general convex entropies for continuous time Markov processes can be described in terms of backward martingales with respect to the tail filtration. The relative entropy is the expected value of a backward submartingale. In the case of (non necessarily reversible) Markov diffusion processes, J. Fontbona and B. Jourdain [65] used Girsanov theory to explicit the Doob-Meyer decomposition of this submartingale. They deduced a stochastic analogue of the well known entropy dissipation formula, which is valid for general convex entropies, including the total variation distance. Under additional regularity assumptions, and using Itô's calculus and ideas of Arnold, Carlen and Ju [42], they obtained a new Bakry-Emery criterion which ensures exponential convergence of the entropy to 0. This criterion is non-intrinsic since it depends on the square root of the diffusion matrix, and cannot be written only in terms of the diffusion matrix itself. They provided examples where the classic Bakry Emery criterion fails, but their non-intrinsic criterion applies without modifying the law of the diffusion process.

With J. Corbetta, A. Alfonsi and B. Jourdain have studied the time derivative of the Wasserstein distance between the marginals of two Markov processes. The Kantorovich duality leads to a natural candidate for this derivative. Up to the sign, it is the sum of the integrals with respect to each of the two marginals of the corresponding generator applied to the corresponding Kantorovich potential. For pure jump processes with bounded intensity of jumps, J. Corbetta, A. Alfonsi and B. Jourdain [15] proved that the evolution of the Wasserstein distance is actually given by this candidate. In dimension one, they showed that this remains true for Piecewise Deterministic Markov Processes. They applied the formula to estimate the exponential decrease rate of the Wasserstein distance between the marginals of two birth and death processes with the same generator in terms of the Wasserstein curvature.

### *3.2.2. Mean-field systems: modeling and control*

**- Mean-field limits of systems of interacting particles.** In [75], B. Jourdain and his former PhD student J. Reygner have studied a mean-field version of rank-based models of equity markets such as the Atlas model introduced by Fernholz in the framework of Stochastic Portfolio Theory. They obtained an asymptotic description of the market when the number of companies grows to infinity. Then, they discussed the long-term capital distribution, recovering the Pareto-like shape of capital distribution curves usually derived from

empirical studies, and providing a new description of the phase transition phenomenon observed by Chatterjee and Pal. They have also studied multitype sticky particle systems which can be obtained as vanishing noise limits of multitype rank-based diffusions (see [74]). Under a uniform strict hyperbolicity assumption on the characteristic fields, they constructed a multitype version of the sticky particle dynamics. In [76], they obtain the optimal rate of convergence as the number of particles grows to infinity of the approximate solutions to the diagonal hyperbolic system based on multitype sticky particles and on easy to compute time discretizations of these dynamics.

In [69], N. Fournier and B. Jourdain are interested in the two-dimensional Keller-Segel partial differential equation. This equation is a model for chemotaxis (and for Newtonian gravitational interaction).

**- Mean field control and Stochastic Differential Games (SDGs).**To handle situations where controls are chosen by several agents who interact in various ways, one may use the theory of Stochastic Differential Games (SDGs). Forward–Backward SDG and stochastic control under Model Uncertainty are studied in [83] by A. Sulem and B. Øksendal. Also of interest are large population games, where each player interacts with the average effect of the others and individually has negligible effect on the overall population. Such an interaction pattern may be modeled by mean field coupling and this leads to the study of mean-field stochastic control and related SDGs. A. Sulem, Y. Hu and B. Øksendal have studied singular mean field control problems and singular mean field two-players stochastic differential games [72]. Both sufficient and necessary conditions for the optimal controls and for the Nash equilibrium are obtained. Under some assumptions, the optimality conditions for singular mean-field control are reduced to a reflected Skorohod problem. Applications to optimal irreversible investments under uncertainty have been investigated. Predictive mean-field equations as a model for prices influenced by beliefs about the future are studied in [85].

### *3.2.3. Stochastic control and optimal stopping (games) under nonlinear expectation*

M.C. Quenez and A. Sulem have studied optimal stopping with nonlinear expectation $\mathcal{E}^g$ induced by a BSDE with jumps with nonlinear driver $g$ and irregular obstacle/payoff (see [82]). In particular, they characterize the value function as the solution of a reflected BSDE. This property is used in [19] to address American option pricing in markets with imperfections. The Markovian case is treated in [59] when the payoff function is continuous.

In [8], M.C. Quenez, A. Sulem and R. Dumitrescu study a combined optimal control/stopping problem under nonlinear expectation $\mathcal{E}^g$ in a Markovian framework when the terminal reward function is only Borelian. In this case, the value function $u$ associated with this problem is irregular in general. They establish a *weak* dynamic programming principle (DPP), from which they derive that the upper and lower semi-continuous envelopes of $u$ are the sub- and super- *viscosity solution* of an associated nonlinear Hamilton-Jacobi-Bellman variational inequality.

The problem of a generalized Dynkin game problem with nonlinear expectation $\mathcal{E}^g$ is addressed in [60]. Under Mokobodzki's condition, we establish the existence of a value function for this game, and characterize this value as the solution of a doubly reflected BSDE. The results of this work are used in [9] to solve the problem of game option pricing in markets with imperfections.
A generalized mixed game problem when the players have two actions: continuous control and stopping is studied in a Markovian framework in [61]. In this work, dynamic programming principles (DPP) are established: a strong DPP is proved in the case of a regular obstacle and a weak one in the irregular case. Using these DPPs, links with parabolic partial integro-differential Hamilton-Jacobi- Bellman variational inequalities with two obstacles are obtained.

With B. Øksendal and C. Fontana, A. Sulem has contributed on the issues of robust utility maximization [84], [85], and relations between information and performance [64].

### *3.2.4. Generalized Malliavin calculus*

Vlad Bally has extended the stochastic differential calculus built by P. Malliavin which allows one to obtain integration by parts and associated regularity probability laws. In collaboration with L. Caramellino (Tor Vegata University, Roma), V. Bally has developed an abstract version of Malliavin calculus based on a splitting

method (see [44]). It concerns random variables with law locally lower bounded by the Lebesgue measure (the so-called Doeblin's condition). Such random variables may be represented as a sum of a "smooth" random variable plus a rest. Based on this smooth part, he achieves a stochastic calculus which is inspired from Malliavin calculus [6]. An interesting application of such a calculus is to prove convergence for irregular test functions (total variation distance and more generally, distribution distance) in some more or less classical frameworks as the Central Limit Theorem, local versions of the CLT and moreover, general stochastic polynomials [48]. An exciting application concerns the number of roots of trigonometric polynomials with random coefficients [49]. Using Kac Rice lemma in this framework one comes back to a multidimensional CLT and employs Edgeworth expansions of order three for irregular test functions in order to study the mean and the variance of the number of roots. Another application concerns U statistics associated to polynomial functions. The techniques of generalized Malliavin calculus developed in [44] are applied in for the approximation of Markov processes (see [52] and [51]). On the other hand, using the classical Malliavin calculus, V. Bally in collaboration with L. Caramellino and P. Pigato studied some subtle phenomena related to diffusion processes, as short time behavior and estimates of tubes probabilities (see [46], [47], [45]).

## 3.3. Numerical Probability

Our project team is very much involved in numerical probability, aiming at pushing numerical methods towards the effective implementation. This numerical orientation is supported by a mathematical expertise which permits a rigorous analysis of the algorithms and provides theoretical support for the study of rates of convergence and the introduction of new tools for the improvement of numerical methods. This activity in the MathRisk team is strongly related to the development of the Premia software.

### *3.3.1. Simulation of stochastic differential equations*

*3.3.1.1. - Weak convergence of the Euler scheme in optimal transport distances.*

With A. Kohatsu-Higa, A. Alfonsi and B. Jourdain [4] have proved using optimal transport tools that the Wasserstein distance between the time marginals of an elliptic SDE and its Euler discretization with $N$ steps is not larger than $C\sqrt{\log(N)}/N$. The logarithmic factor may is removed when the uniform time-grid is replaced by a grid still counting $N$ points but refined near the origin of times.

*3.3.1.2. - Strong convergence properties of the Ninomiya Victoir scheme and multilevel Monte-Carlo estimators.*

With their former PhD student, A. Al Gerbi, E. Clément and B. Jourdain [1] have proved strong convergence with order $1/2$ of the Ninomiya-Victoir scheme which is known to exhibit order 2 of weak convergence [81]. This study was aimed at analysing the use of this scheme either at each level or only at the finest level of a multilevel Monte Carlo estimator : indeed, the variance of a multilevel Monte Carlo estimator is related to the strong error between the two schemes used in the coarse and fine grids at each level. In [14], they proved that the order of strong convergence of the crude Ninomiya Victoir scheme is improved to 1 when the vector fields corresponding to each Brownian coordinate in the SDE commute, and in [34], they studied the error introduced by discretizing the ordinary differential equations involved in the Ninomiya-Victoir scheme.

*3.3.1.3. - Non-asymptotic error bounds for the multilevel Monte Carlo Euler method.*

A. Kebaier and B. Jourdain are interested in deriving non-asymptotic error bounds for the multilevel Monte Carlo method. As a first step, they dealt in [73] with the explicit Euler discretization of stochastic differential equations with a constant diffusion coefficient. They obtained Gaussian-type concentration. To do so, they used the Clark-Ocone representation formula and derived bounds for the moment generating functions of the squared difference between a crude Euler scheme and a finer one and of the squared difference of their Malliavin derivatives. The estimation of such differences is much more complicated than the one of a single Euler scheme contribution and explains why they suppose the diffusion coefficient to be constant. This assumption ensures boundedness of the Malliavin derivatives of both the SDE and its Euler scheme.

*3.3.1.4. - Computation of sensibilities of integrals with respect to the invariant measure.*

In [43], R. Assaraf, B. Jourdain, T. Lelièvre and R. Roux considered the solution to a stochastic differential equation with constant diffusion coefficient and with a drift function which depends smoothly on some real parameter $\lambda$, and admitting a unique invariant measure for any value of $\lambda$ around $\lambda = 0$. Their aim was to compute the derivative with respect to $\lambda$ of averages with respect to the invariant measure, at $\lambda = 0$. They analyzed a numerical method which consists in simulating the process at $\lambda = 0$ together with its derivative with respect to $\lambda$ on a long time horizon. They gave sufficient conditions implying uniform-in-time square integrability of this derivative. This allows in particular to compute efficiently the derivative with respect to $\lambda$ of the mean of an observable through Monte Carlo simulations.

*3.3.1.5. - Approximation of doubly reflected Backward stochastic differential equations.*

R. Dumitrescu and C. Labart have studied the discrete time approximation scheme for the solution of a doubly reflected Backward Stochastic Differential Equation with jumps, driven by a Brownian motion and an independent compensated Poisson process [58], [57].

*3.3.1.6. - Parametrix methods.*

V. Bally and A. Kohatsu-Higa have recently proposed an unbiased estimator based on the parametrix method to compute expectations of functions of a given SDE ( [50]). This method is very general, and A. Alfonsi, A. Kohastu-Higa and M. Hayashi [37] have applied it to the case of one-dimensional reflected diffusions. In this case, the estimator can be obtained explicitly by using the scheme of Lépingle [78] and is quite simple to implement. It is compared to other simulation methods for reflected SDEs.

### *3.3.2. Estimation of the parameters of a Wishart process*

A. Alfonsi, A. Kebaier and C. Rey [38] have computed the Maximum Likelihood Estimator for the Wishart process and studied its convergence in the ergodic and in some non ergodic cases. In the ergodic case, which is the most relevant for applications, they obtain the standard square-root convergence. In the non ergodic case, the analysis rely on refined results for the Laplace transform of Wishart processes, which are of independent interest.

### *3.3.3. Optimal stopping and American options*

In joint work with A. Bouselmi, D. Lamberton studied the asymptotic behavior of the exercise boundary near maturity for American put options in exponential Lévy models. In [7], they deal with jump-diffusion models, and establish that, in some cases, the behavior differs from the classical Black and Scholes setting. D. Lamberton has also worked on the binomial approximation of the American put. The conjectured rate of convergence is $O(1/n)$ where $n$ is the number of time periods. He was able to derive a $O((\ln n)^{\alpha}/n)$ bound, where the exponent $\alpha$ is related to the asymptotic behavior of the exercise boundary near maturity.

# 4. Application Domains

## 4.1. Financial Mathematics, Insurance

The domains of application are quantitative finance and insurance with emphasis on risk modeling and control. In particular, Mathrisk focuses on dependence modeling, systemic risk, market microstructure modeling and risk measures.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

The project team Mathrisk has been evaluated in March. The report was very positive.

# 6. New Software and Platforms

## 6.1. PREMIA

KEYWORDS: Financial products - Computational finance - Option pricing
SCIENTIFIC DESCRIPTION: The Premia project keeps track of the most recent advances in the field of computational finance in a well-documented way. It focuses on the implementation of numerical analysis techniques for both probabilistic and deterministic numerical methods. An important feature of the platform Premia is the detailed documentation which provides extended references in option pricing.

Premia is thus a powerful tool to assist Research and Development professional teams in their day-to-day duty. It is also a useful support for academics who wish to perform tests on new algorithms or pricing methods without starting from scratch.

Besides being a single entry point for accessible overviews and basic implementations of various numerical methods, the aim of the Premia project is: 1 - to be a powerful testing platform for comparing different numerical methods between each other, 2 - to build a link between professional financial teams and academic researchers, 3 - to provide a useful teaching support for Master and PhD students in mathematical finance.
FUNCTIONAL DESCRIPTION: Premia is a software designed for option pricing, hedging and financial model calibration.

- Participants: Agnes Sulem, Antonino Zanette, Aurélien Alfonsi, Benjamin Jourdain, Jérôme Lelong and Bernard Lapeyre
- Partners: Inria - Ecole des Ponts ParisTech - Université Paris-Est
- Contact: Agnes Sulem
- URL: http://www.premia.fr

## 6.2. Platforms

### *6.2.1. Development of the quantitative platform Premia in 2018*

Premia 20 has been delivered to the Consortium Premia on March 12th. It contains the following new algorithms :

*6.2.1.1. Optimal Trade Execution, Risk Management, Insurance*

- Optimal Execution Under Jump Models For Uncertain Price Impact. S.Moazeni, T.F.Coleman, Y.Li
  *The Journal of Computational Finance. Vol. 18, Issue 3, 2015.*
- Nested Monte Carlo for Risk Margin computation. L.A. Abbas-Turki, S.Crepey, B.Diallo.
- Efficient Estimation of Sensitivities for Counterparty Credit Risk with the Finite Difference Monte-Carlo Method. C. S.L. de Graaf, D.Kandhai, P.M.A.Sloot.
  *The Journal of Computational Finance, Volume 21, Issue 1, 2017.*
- Nested Simulation in Portfolio Risk Measurement. M.B.Gordy, S.Juneja
  *Management Science, Vol 56, Issue 10, 2010*
- Spectral methods for the calculation of risk measures for variable annuity guaranteed benefits. R. Feng, H.W. Volkmer
  *ASTIN Bull., 44(3), 2014*
- Fast computation of risk measures for variable annuities with additional earnings by conditional moment matching. N. Privault X.Wei
  *ASTIN Bull., 48(1):171–196, 2018.*

*6.2.1.2. Equity Derivatives*

- Pricing under Rough volatility. C. Bayer, P.Friz, J. Gatheral
  *Quantitative Finance, Vol. 16, No. 6, 887-904, 2016.*

- Hybrid scheme for Brownian semistationary processes. M. Bennedsen, A. Lunde, M.S.Pakkanen
  *Finance and Stochastics 21(4), 931–965, 2017.*
- Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation. M. B. Giles and L. Szpruch
  *The Annals of Applied Probability, Vol. 24, No. 4, 2014*
- Fourier transform algorithms for pricing and hedging discretely sampled exotic variance products and volatility derivatives under additive processes. W. Zheng and Y. K. Kwok
  *The Journal of Computational Finance, Volume 18, Issue 2, 2014.*
- Efficient Solution of Backward Jump-Diffusion PIDEs with Splitting and Matrix Exponentials. A.Iktin
  *The Journal of Computational Finance, Volume 19, Issue 3, 2016*
- High-Order Splitting Methods for Forward PDEs and PIDEs. A.Iktin
  *International Journal of Theoretical and Applied Finance, 18(5), 2015*
- Pricing Bullet option on local volatility model using GPU L.A. Abbas-Turki
- Pricing Bermudan Options via Multilevel Approximation Methods. D. Belomestny, F. Dickmann, T.Nagapetyan.
  *Siam J. Financial Math., Volume 6, 2015.*
- Pricing CIR yield options by conditional moment matching. A. Prayoga N. Privault
  *Asia-Pacific Financial Markets, 24:19–38, 2017*


We benefit from the help of the engineer Pierre-Guillaume Raverdy.

# 7. New Results

## 7.1. Risk management in finance and insurance

### *7.1.1. Control of systemic risk in a dynamic framework*

Interconnected systems are subject to contagion in time of distress. Recent effort has been dedicated to understanding the relation between network topology and the scope of distress propagation. It is critical to recognize that connectivity is a result of an optimization problem of agents, who derive benefits from connections and view the associated contagion risk as a cost. In our previous works on the control of contagion in financial systems (see e.g. [80], [41], [5]), a central party, for example a regulator or government, seeks to minimize contagion. In [54], in contrast, the financial institutions themselves are the decision makers, and their decision is made before the shock, with a rational expectation on the way the cascade will evolve following the shock. We are extending these studies in a *dynamic* framework by allowing a recovery feature in the financial system during the cascade process, captured by introducing certain extent of growth of the banks' assets between each round of contagion.

### *7.1.2. Option pricing in financial markets with imperfections and default*

A. Sulem, M.C. Quenez and R. Dumitrescu have studied robust pricing in an imperfect financial market with default. In this setting, the pricing system is expressed as a nonlinear g-expectation $\mathcal{E}^g$ induced by a nonlinear BSDE with nonlinear driver $g$ and default jump (see [24]). The case of American options in this market model is treated in [19]. The incomplete market case is under study.

### *7.1.3. American options*

With Giulia Terenzi, D. Lamberton has been been working on American options in Heston's model. They have some results about existence and uniqueness for the associated variational inequality, in suitable weighted Sobolev spaces (see Feehan and co-authors for recent results on elliptic problems). Their paper "Variational formulation of American option prices in the Heston model" [32] is now in minor revision for *SIAM Journal on Financial Mathematics*.

They also have some results on monotonicity and regularity properties of the price function.

D. Lamberton has also a paper on the binomial approximation of the American put, in which a new bound for the rate of convergence of the binomial approximation of the Black-Scholes American put price is derived [32].

Optimal stopping problems involving the maximum of a diffusion is currently under investigation. Partial results obtained by D. Lamberton and M. Zervos) enable them to treat reward functions with little regularity.

### *7.1.4. Monte-Carlo methods for the computation of the Solvency Capital Requirement (SCR) in Insurance*

A. Alfonsi has obtained a grant from AXA Foundation on a Joint Research Initiative with a team of AXA France working on the strategic asset allocation. This team has to make recommendations on the investment over some assets classes as, for example, equity, real estate or bonds. In order to do that, each side of the balance sheet (assets and liabilities) is modeled in order to take into account their own dynamics but also their interactions. Given that the insurance products are long time contracts, the projections of the company's margins have to be done considering long maturities. When doing simulations to assess investment policies, it is necessary to take into account the SCR which is the amount of cash that has to be settled to manage the portfolio. Typically, the computation of the future values of the SCR involve expectations under conditional laws, which is greedy in computation time. The goal of this project is to develop efficient Monte-Carlo methods to compute the SCR for long investment strategies. A. Cherchali has started his PhD thesis in September 2017 on this topic.

A. Alfonsi and A. Cherchali are developing a model of the ALM management of insurance companies that takes into account the regulatory constraints on life-insurance. We are testing this model. The purpose is then to use this model to develop Monte-Carlo methods to approximate the SCR (Solvency Capital Requirement).

## 7.2. Optimal transport and applications

### *7.2.1. Martingale Optimal Transport.*

B. Jourdain and W. Margheriti exhibit a new family of martingale couplings between two one-dimensional probability measures $\mu$ and $\nu$ in the convex order. This family is parametrised by two dimensional probability measures on the unit square with respective marginal densities proportional to the positive and negative parts of the difference between the quantile functions of $\mu$ and $\nu$. It contains the inverse transform martingale coupling which is explicit in terms of the cumulative distribution functions of these marginal densities. The integral of $|x-y|$ with respect to each of these couplings is smaller than twice the $W^1$ distance between $\mu$ and $\nu$. When the comonotoneous coupling between $\mu$ and $\nu$ is given by a map $T$, the elements of the family minimize $\int_{\mathbf{R}} |y - T(x)| M(dx, dy)$ among all martingale couplings $M$ between $\mu$ and $\nu$. When $\mu$ and $\nu$ are in the decreasing (resp. increasing) convex order, the construction can be generalized to exhibit super (resp. sub) martingale couplings.

A. Alfonsi and B. Jourdain show that any optimal coupling for the quadratic Wasserstein distance $W_2^2(\mu,\nu)$ between two probability measures $\mu$ and $\nu$ with finite second order moments on $\mathbf{R}^d$ is the composition of a martingale coupling with an optimal transport map $\mathcal{T}$. They check the existence of optimal couplings in which this map gives the unique optimal coupling between $\mu$ and $\mathcal{T}\#\mu$. Next, they prove that $\sigma \mapsto W_2^2(\sigma,\nu)$ is differentiable at $\mu$ in both Lions and the geometric senses iff there is a unique optimal coupling between $\mu$ and $\nu$ and this coupling is given by a map.

### *7.2.2. Numerical methods for optimal transport.*

Optimal transport problems have got a recent attention in many different fields including physics, quantum chemistry and finance, where Martingale Optimal Transport problems allow to quantify the model risk. In practice, few numerical methods exist to approximate the optimal coupling measure and/or the optimal transport. In particular, to deal with large dimensions or with the optimal transport problems with many marginal laws, a natural direction is to develop Monte-Carlo methods.

A. Alfonsi, V. Ehrlacher (CERMICS, Inria Project-team MATHERIALS), D. Lombardi (Inria Project-team Reo) and R. Coyaud (PhD student of A. Alfonsi) are working on numerical approximations of the optimal transport between two (or more) probability measures.

## 7.3. Optimal Control of Mean field (S)PDEs

With Rui Chen and R. Dumitrescu, A. Sulem has studied mean-field Backward SDEs driven by a Brownian motion and an independant Poisson random measure and its interpretation in terms of global risk measures. Dual representation has been provided in the convex case. Optimal stopping for these BSDEs and links with reflected mean-field BSDEs has also been investigated.

A. Sulem, R. Dumitrescu and B. Øksendal have studied optimal control for mean-field stochastic **partial** differential equations (stochastic evolution equations) driven by a Brownian motion and an independent Poisson random measure, in the case of *partial information* control [20]. One important novelty is the introduction of *general mean-field* operators, acting on both the controlled state process and the control process. A sufficient and a necessary maximum principle for this type of control is formulated. Existence and uniqueness of the solution of such general forward and backward mean-field stochastic partial differential equations are proved. These results have been applied to find the explicit optimal control for an optimal harvesting problem.

## 7.4. Analysis of probabilistic numerical methods

### *7.4.1. Particles approximation of mean-field SDEs*

O. Bencheikh and Benjamin Jourdain have proved that the weak error between a stochastic differential equation with nonlinearity in the sense of McKean given by moments and its approximation by the Euler discretization with time-step $h$ of a system of $N$ interacting particles is $\mathcal{O}(N^{-1}+h)$. Numerical experiments confirm this behaviour and show that it extends to more general mean-field interaction.

### *7.4.2. Approximation of Markov processes*

V. Bally worked on general approximation schemes in total variation distance for diffusion processes in collaboration with his former Phd student Clément Rey [52] This work includes high order schemes as Victoir-Ninomya for example. Further development in this direction is under study in collaboration with A. Alfonsi. Moreover, in collaboration with his former Phd student V. Rabiet and with D. Goreac (University Paris Est Marne la Vallée), V. Bally is studying approximations schemes for Piecewise Deterministic Markov Processes (see [17], [51]). In this framework the goal is to replace small jumps by a Brownien component - such a procedure is popular for "usual" jump equations, but the estimate of the error in the case of PDMP's is much more delicate. A significant example is the Bolzmann equation [28].

### *7.4.3. High order approximation for diffusion processes*

A. Alfonsi and V. Bally are working on a generic method to achieve any weak order of convergence for approximating SDEs.

### *7.4.4. Adaptive MCMC methods*

The Self-Healing Umbrella Sampling (SHUS) algorithm is an adaptive biasing algorithm which has been proposed in order to efficiently sample a multimodal probability measure.

In [21], G. Fort, B. Jourdain, T. Lelièvre and G. Stoltz extend previous works [68], [66], [67] and study a larger class of algorithms where the target distribution is biased using only a fraction of the free energy and which includes a discrete version of well-tempered metadynamics.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

- Consortium PREMIA, Natixis - Inria
- Consortium PREMIA, Crédit Agricole Corporate Investment Bank (CA - CIB ) - Inria
- Chair X-ENPC-UPMC-Société Générale "Financial Risks" of the Risk fondation : A. Alfonsi, B. Jourdain, B. Lapeyre
- AXA Joint Research Initiative on Numerical methods for the ALM, from September 2017 to August 2020. PhD grant of Adel Cherchali, Supervisor: A. Alfonsi.
- CIFRE agreement Milliman company/Ecole des Ponts (http://fr.milliman.com),

  PhD thesis of Sophian Mehalla (started November 2017) on "Interest rate risk modeling for insurance companies", Supervisor: Bernard Lapeyre.
- Collaboration with IRT Systemx

  PhD grant of Adrien Touboul (started November 2017) on "Uncertainty computation in a graph of physical simulations", Supervisors: Bernard Lapeyre and Julien Reygner.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

- ANR Cosmos 2015-2018, Participant: B. Jourdain ; Partners : Ecole des Ponts, Telecom, Inria Rennes and IBPC
- Labex Bezout
  http://bezout.univ-paris-est.fr

### *9.1.1. Competitivity Clusters*

**Pôle Finance Innovation**

## 9.2. International Initiatives

### *9.2.1. Informal International Partners*

- Center of Excellence program in Mathematics and Life Sciences at the Department of Mathematics, University of Oslo, Norway, (B. Øksendal).
- Kings College, London (R. Dumitrescu)
- Department of Mathematics, University of Manchester (Tusheng Zhang, currently in charge of an EU-ITN program on BSDEs and Applications).
- Kensas University (Yaozhong Hu)
- Cornell University, ORIE department (Andreea Minca)
- Mannheim University (Alexander Schied, Chair of Mathematics in Business and Economics, Department of Mathematics)
- Roma Tor Vergata University (Lucia Caramellino)
- Ritsumeikan University (A. Kohatsu-Higa).

## 9.3. International Research Visitors

### *9.3.1. Visits of International Scientists*

- Oleg Kudryavtsev, Rostov University (Russia)
- B. Stemper (Weierstrass Institute Berlin)
- A. Kohatsu Higa (Ritsumeikan University)

### *9.3.2. Internships*

Oussama Bellalah, Inria, May-August

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. Member of the Organizing Committees*

- A. Alfonsi:

  Co-organizer of the working group seminar of MathRisk “Méthodes stochastiques et finance”.

- A. Sulem

  Co-organizer of the seminar Inria-MathRisk /Université Paris 7 LPMA “Numerical probability and mathematical finance”. https://www.lpsm.paris/mathfipronum/gt

### *10.1.2. Journal*

*10.1.2.1. Member of the Editorial Boards*

- B. Jourdain

  Associate editor of

  - ESAIM : Proceedings and Surveys
  - Stochastic Processes and their Applications (SPA)

- D. Lamberton

  Associate editor of

  - Mathematical Finance,
  - Associate editor of ESAIM Probability & Statistics

- A. Sulem

  Associate editor of

  - Journal of Mathematical Analysis and Applications (JMAA)
  - International Journal of Stochastic Analysis (IJSA)
  - SIAM Journal on Financial Mathematics (SIFIN)

*10.1.2.2. Reviewer - Reviewing Activities*

- B. Jourdain : Reviewer for *Mathematical Reviews*
- A. Sulem: Reviewer for *Mathematical Reviews*

### *10.1.3. Invited Talks*

- A. Alfonsi
  - 15th of December 2017: "Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems." Séminaire Bachelier, Paris.
  - 12th of June 2018: "Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems." Conference on Stochastic modeling and financial applications, Verona.
  - 29th and 30th of August 2018: "Introduction to affine processes". Lecture given at the 11th European Summer School in Financial Mathematics, Palaiseau.
  - 29th of October 2018: "Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems." International Conference on Control, Games and Stochastic Analysis, Hammamet.

  - 7th of December 2018: "Approximation de mesures de probabilité dans l'ordre convexe par projections pour la distance de Wasserstein." Séminaire de Mathématiques Appliquées, Collège de France.
- V. Bally
  - Conference SPA (Stochastic Processes and their Applications): "Abstract Malliavin calculus and invariance principles", 11-15 June 2018, Gothenburg, Sweden.
  - Workshop on Analytical Aspects of Stochastic Systems: "Transfer of regularity for Markov semigroups", Växjö, Sweden, June 6-8, 2018.
  - Workshop Recent Advances in Random Processes - Conference in honor of Paolo Baldi's 70th aniversary. Talk: Malliavin Calculus and Invariance Principles"
  - Workshop on Asymptotic expansions and Malliavin calculus 15-16 November 2018, Institut Henri Poincaré. Talk: Malliavin Calculus and Invariance Principles"
- B. Jourdain

  - Inria Mathrisk/LPSM university Paris Diderot seminar, 20 December 2018 : Differentiability of the squared quadratic Wasserstein distance

  - 1st Moscow-UK workshop on stochastic analysis : Wasserstein calculus and related topics, ICMS Edinburgh 19-23 November 2018 : Lifted and geometric differentiability of the squared quadratic Wasserstein distance

  - Seminar of the chair Financial Risks, IHP, 5 October 2018 : A new family of one-dimensional martingales couplings

  - Populations : Interaction and Evolution, IHP, 10-14 September 2018 : Existence of a calibrated regime-switching local volatility model

  - Journées MAS 2018, Dijon, 29-31 August : plenary talk entitled Sampling of probability measures in the convex order and approximation of martingale optimal transport problems

  - 10th world congress of the Bachelier finance society, Dublin, 16-20 July 2018 : Sampling of probability measures in the convex order and approximation of martingale optimal transport problems

  - MCQMC2018, Rennes, 2-6 July 2018 : Sampling of probability measures in the convex order and approximation of martingale optimal transport problems

  - Bachelier course : Systems of rank-based diffusions with mean-field interaction, 4 hours, 23 and 30 March 2018
- D. Lamberton

  Invited speaker: Symposium on optimal stopping, June 25-29 2018, Rice University, Houston. (USA)
- A. Sulem
  - Conference SPA (Stochastic Processes and their Applications): "Stochastic Optimal Control Under Partial Observations", 11-15 June 2018, Gothenburg, Sweden.

### *10.1.4. Research Administration*

- A. Alfonsi

  - Deputy director of CERMICS laboratory

  - In charge of the Master "Finance and Application" at the Ecole des Ponts.
- D. Lamberton

  Vice-president for research at Université Paris-Est Marne-la-Vallée
- B. Jourdain

  - Head of the doctoral school MSTIC, University Paris-Est

- A. Sulem

  - Member of the Committee for scientific positions (Commission des emplois scientifiques), Inria Paris

  - Corresponding member of the comité opérationel d'évaluation des risques légaux et éthiques (COERLE) at Inria Paris research center

  - Member of the Committee for Inria international Chairs

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Licence :

- A. Alfonsi: 'Probabilités", first year course at Ecole des Ponts.

- V. Bally : "Analyse Hilbertienne", Course L3, UPEMLV

Master :

- A. Alfonsi:
  - "Données Haute Fréquence en finance",Master lectures at UPEMLV.
  - 'Traitement des données de marché : aspects statistiques et calibration", Master lectures at UPEMLV.
  - "Mesures de risque", Master course of UPEMLV and Paris VI.
  - Professeur chargé de cours at Ecole Polytechnique.
- V. Bally
  - Course on "Taux d'Intêret", M2 Finance, UPEMLV
  - Course on "Calcul de Malliavin et applications en finance", M2 Finance, UPEMLV
  - Course on "Analyse du risque", M2 Actuariat, UPEMLV
  - Course on "Processus Stochastiques", M2 Recherche, UPEMLV
- B. Jourdain
  - Course on "Mont-Carlo Markov chain methods and particle algorithms", Research Master Probabilités et Modèles Aléatoires, Sorbonne Université
  - B. Jourdain: course "Mathematical finance", M1, ENPC
- B. Jourdain, B. Lapeyre
  - Course "Monte-Carlo methods", 3rd year ENPC and Master Recherche Mathématiques et Application, University of Marne-la-Vallée
- J.-F. Delmas, B.Jourdain
  - course "Jump processes with applications to energy markets", 3rd year ENPC and Master Recherche Mathématiques et Application, University of Marne-la-Vallée
- B. Lapeyre
  - Monte-Carlo methods in quantitative finance, Master of Mathematics, University of Luxembourg,
- D. Lamberton
  - Calcul stochastique pour la finance, master 1 course, Université Paris-Est Marne-la-Vallée
- A. Sulem
  - "Finite difference for PDEs in Finance", Master 2 MASEF, Université Paris IX-Dauphine, Département Mathématiques et Informatique de la Décision et des Organisations (MIDO), 27 h.

- "PDE methods in Finance", Master of Mathematics, University of Luxembourg, 22 h lectures and responsible of the module "Numerical Methods in Finance".

### *10.2.2. Supervision*

PhD Alexandre Zhou, "Theoretical and numerical study of problems nonlinear in the sense of McKean in finance", Ecole des Ponts, defended on October 17th 2018, supervised by B.Jourdain.

PhD Giulia Terenzi, "American options in complex financial models", université Marne la Vallée, defended on December 17th 2018, supervised by D. Lamberton and Lucia Caramellino, University Tor Vergata, Rome.

PhD Marouan Iben Taarit , " On CVA and XVA computations ", "Valorisation des ajustements Xva : de l'exposition espérée aux risques adverses de corrélation", CIFRE Natixis/ENPC, defended on January 8th, ENPC, Supervisor: Bernard Lapeyre.

PhD in progress :

- Anas Bentaleb (started February 2018) : Mathematical techniques for expected exposure evaluation, Supervisor: B. Lapeyre.

- Adel Cherchali, "Numerical methods for the ALM", funded by Fondation AXA, starting from September 2017, Supervisor: A. Alfonsi

- Rafaël Coyaud, "Deterministic ans stochastic numerical methods for multimarginal and martingale constraint optimal transport problems", starting from October 2017, Supervisor: A. Alfonsi

- Rui Chen (Fondation Sciences Mathématiques de Paris grant), "Stochastic Control of mean field systems and applications to systemic risk, from September 2014, Université Paris-Dauphine, Supervisor: A. Sulem.

- Sophian Mehalla (started November 2017), CIFRE agreement Milliman company/Ecole des Ponts (http://fr.milliman.com, Supervisor: B. Lapeyre

- Oumaima Bencheikh (started November 2017) "Acceleration of probabilistic particle methods", Supervisor: B. Jourdain

- Ezechiel Kahn (started September 2018) "Functional inequalities for random matrices models", supervised by B. Jourdain and D. Chafai

- William Margheriti (started January 2018) "Numerical methods for martingale optimal transport problems", supervised by J.-F. Delmas and B. Jourdain

### *10.2.3. Juries*

B. Jourdain :

Jury and report on

- PhD of Hadrien De March, defended on June 29, university Paris Saclay
- PhD of David Krief, defended on September 27, University Paris Diderot

A. Sulem

- PhD of David Krief, defended on September 27, University Paris Diderot (Chair of the Committee)
- PhD of Xiao Wei, November 27, University Paris Diderot
- PhD Hadjer Moussaoui, December 14, Université de Toulon
- HdR of Thomas Lim, ENSIIE, December 4, Université Evry Val d'Essonne
- PEDR CNRS September 2018
- PRIX Inria : Grand Prix Inria - Académie des Sciences; - Prix Jeune Chercheur - Académie des Sciences ; Prix Innovation - Dassault Système (Spring 2018)

# 11. Bibliography

## Major publications by the team in recent years

[1] A. AL GERBI, B. JOURDAIN, E. CLÉMENT.*Ninomiya-Victoir scheme: strong convergence, antithetic version and application to multilevel estimators*, in "Monte Carlo Method and Applications", July 2016, vol. 22, n° 3, p. 197-228, https://arxiv.org/abs/1508.06492 , https://hal-enpc.archives-ouvertes.fr/hal-01188675

[2] A. ALFONSI.*Affine diffusions and related processes: simulation, theory and applications*, Bocconi and Springer Series, Mathematics statistics, finance and economics, Springer, 2015

[3] A. ALFONSI, P. BLANC.*Dynamic optimal execution in a mixed-market-impact Hawkes price model*, in "Finance and Stochastics", January 2016, https://arxiv.org/abs/1404.0648 [*DOI :* 10.1007/s00780-015-0282-y], https://hal-enpc.archives-ouvertes.fr/hal-00971369

[4] A. ALFONSI, B. JOURDAIN, A. KOHATSU-HIGA.*Optimal transport bounds between the time-marginals of a multidimensional diffusion and its Euler scheme*, in "Electronic Journal of Probability", 2015, https://arxiv.org/abs/1405.7007 , https://hal-enpc.archives-ouvertes.fr/hal-00997301

[5] H. AMINI, A. MINCA, A. SULEM.*Control of interbank contagion under partial information*, in "SIAM Journal on Financial Mathematics", December 2015, vol. 6, n° 1, 24, https://hal.inria.fr/hal-01027540

[6] V. BALLY, L. CARAMELLINO.*Convergence and regularity of probability laws by using an interpolation method*, in "Annals of Probability", 2017, vol. 45, n° 2, p. 1110–1159, https://hal-upec-upem.archives-ouvertes.fr/hal-01109276

[7] A. BOUSELMI, D. LAMBERTON.*The critical price of the American put near maturity in the jump diffusion model*, in "SIAM Journal on Financial Mathematics", May 2016, vol. 7, n° 1, p. 236–272, https://arxiv.org/abs/1406.6615 [*DOI :* 10.1137/140965910], https://hal-upec-upem.archives-ouvertes.fr/hal-00979936

[8] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM.*A Weak Dynamic Programming Principle for Combined Optimal Stopping/Stochastic Control with $E^f$-Expectations*, in "SIAM Journal on Control and Optimization", 2016, vol. 54, n° 4, p. 2090-2115 [*DOI :* 10.1137/15M1027012], https://hal.inria.fr/hal-01370425

[9] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM.*Game Options in an Imperfect Market with Default*, in "SIAM Journal on Financial Mathematics", January 2017, vol. 8, n° 1, p. 532 - 559 [*DOI :* 10.1137/16M1109102], https://hal.inria.fr/hal-01614758

[10] B. ØKSENDAL, A. SULEM.*Applied Stochastic Control of Jump Diffusions*, Universitext, Second Edition, Springer, Berlin, Heidelberg, New York, 257 pages 2007, 3rd edition to appear in 2019

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] M. IBEN TAARIT.*Pricing of XVA adjustments : from expected exposures to wrong-way risks*, Université Paris-Est, January 2018, https://pastel.archives-ouvertes.fr/tel-01939269

[12] G. TERENZI.*Option prices in stochastic volatility models*, Université Paris Est - Marne-la-Vallée ; Università di Roma Tor Vergata, December 2018, https://hal.archives-ouvertes.fr/tel-01961071

[13] A. ZHOU.*Theoretical and numerical study of problems nonlinear in the sense of McKean in finance*, Université Paris Est, École des Ponts Paris Tech, 6-8 avenue Blaise Pascal, 77455 Marne La Vallée, October 2018, https://hal.inria.fr/tel-01957638

### Articles in International Peer-Reviewed Journal

[14] A. AL GERBI, B. JOURDAIN, E. CLÉMENT.*Asymptotics for the normalized error of the Ninomiya-Victoir scheme*, in "Stochastic Processes and their Applications", 2018, vol. 128, n$^{o}$ 6, p. 1889-1928, https://arxiv.org/abs/1601.05268 , https://hal-enpc.archives-ouvertes.fr/hal-01259915

[15] A. ALFONSI, J. CORBETTA, B. JOURDAIN.*Evolution of the Wasserstein distance between the marginals of two Markov processes*, in "Bernoulli", 2018, vol. 24, n$^{o}$ 4A, p. 2461-2498, https://arxiv.org/abs/1606.02994 , https://hal.archives-ouvertes.fr/hal-01390887

[16] V. BALLY, L. CARAMELLINO, G. POLY.*Convergence in distribution norms in the CLT for non identical distributed random variables*, in "Electronic Journal of Probability", 2018, vol. 23, 51, paper 45 https://arxiv.org/abs/1606.01629 [*DOI :* 10.1214/18-EJP174], https://hal-upec-upem.archives-ouvertes.fr/hal-01413548

[17] V. BALLY, D. GOREAC, V. RABIET.*Regularity and Stability for the Semigroup of Jump Diffusions with State-Dependent Intensity*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", August 2018, vol. 28, n$^{o}$ 5, p. 3028 - 3074, https://arxiv.org/abs/1707.02713 [*DOI :* 10.1214/18-AAP1382], https://hal.archives-ouvertes.fr/hal-01558741

[18] O. BENCHEIKH, B. JOURDAIN.*Bias behaviour and antithetic sampling in mean-field particle approximations of SDEs nonlinear in the sense of McKean*, in "ESAIM: Proceedings and Surveys", 2018, https://arxiv.org/abs/1809.06838 - 14 pages, https://hal.archives-ouvertes.fr/hal-01877002

[19] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM.*American Options in an Imperfect Complete Market with Default*, in "ESAIM: Proceedings and Surveys", 2018, p. 93–110 [*DOI :* 10.1051/PROC/201864093], https://hal.inria.fr/hal-01614741

[20] R. DUMITRESCU, B. ØKSENDAL, A. SULEM.*Stochastic control for mean-field Stochastic Partial Differential Equations with jumps*, in "Journal of Optimization Theory and Applications", March 2018, p. 559-584 [*DOI :* 10.1007/s10957-018-1243-3], https://hal.inria.fr/hal-01527225

[21] G. FORT, B. JOURDAIN, T. LELIÈVRE, G. STOLTZ.*Convergence and efficiency of adaptive importance sampling techniques with partial biasing*, in "Journal of Statistical Physics", March 2018, vol. 171, n$^{o}$ 2, p. 220–268, https://arxiv.org/abs/1610.09194 [*DOI :* 10.1007/s10955-018-1992-2], https://hal.archives-ouvertes.fr/hal-01389996

[22] L. GOUDENÈGE, A. MOLENT, A. ZANETTE.*Pricing and hedging GMWB in the Heston and in the Black–Scholes with stochastic interest rate models*, in "Computational Management Science", March 2018 [*DOI :* 10.1007/s10287-018-0304-2], https://hal.archives-ouvertes.fr/hal-01940715

[23] D. LAMBERTON.*On the binomial approximation of the American put*, in "Applied Mathematics and Optimization", 2018, https://arxiv.org/abs/1802.05614 , https://hal-upec-upem.archives-ouvertes.fr/hal-01709298

### Scientific Books (or Scientific Book chapters)

[24] R. DUMITRESCU, M. GRIGOROVA, M.-C. QUENEZ, A. SULEM.*BSDEs with default jump*, in "Computation and Combinatorics in Dynamics, Stochastics and Control - The Abel Symposium, Rosendal, Norway August 2016", E. CELLEDONI, G. D. NUNNO, K. EBRAHIMI-FARD, H. MUNTHE-KAAS (editors), The Abel Symposia book series, Springer, 2018, vol. 13 [*DOI :* 10.1007/978-3-030-01593-0], https://hal.inria.fr/hal-01799335

### Other Publications

[25] A. ALFONSI, J. CORBETTA, B. JOURDAIN.*Sampling of one-dimensional probability measures in the convex order and computation of robust option price bounds*, December 2018, This paper is an updated version of a part of the paper https://hal.archives-ouvertes.fr/hal-01589581 (or https://arxiv.org/pdf/1709.05287.pdf ), https://hal-enpc.archives-ouvertes.fr/hal-01963507

[26] A. ALFONSI, B. JOURDAIN.*Lifted and geometric differentiability of the squared quadratic Wasserstein distance*, November 2018, working paper or preprint, https://hal-enpc.archives-ouvertes.fr/hal-01934705

[27] A. ALFONSI, D. KRIEF, P. TANKOV.*Long-time large deviations for the multi-asset Wishart stochastic volatility model and option pricing*, December 2018, https://arxiv.org/abs/1806.06883 - working paper or preprint, https://hal-enpc.archives-ouvertes.fr/hal-01949485

[28] V. BALLY.*Upper bounds for the function solution of the homogenuous 2D Boltzmann equation with hard potential*, May 2018, https://arxiv.org/abs/1710.00695 - working paper or preprint, https://hal-upec-upem.archives-ouvertes.fr/hal-01593131

[29] V. BALLY, L. CARAMELLINO.*Convergence and regularity of probability laws by using an interpolation method*, January 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01677407

[30] L. GOUDENÈGE, A. MOLENT, A. ZANETTE.*Computing Credit Valuation Adjustment solving coupled PIDEs in the Bates model*, September 2018, https://arxiv.org/abs/1809.05328 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01873346

[31] B. JOURDAIN, W. MARGHERITI.*A new family of one dimensional martingale couplings*, September 2018, https://arxiv.org/abs/1808.01390 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01876809

[32] D. LAMBERTON, G. TERENZI.*Variational formulation of American option prices in the Heston Model*, December 2018, https://arxiv.org/abs/1711.11311 - working paper or preprint, https://hal-upec-upem.archives-ouvertes.fr/hal-01649496

## References in notes

[33] A. AHDIDA, A. ALFONSI, E. PALIDDA.*Smile with the Gaussian term structure model*, in "Journal of Computational Finance", 2017, https://arxiv.org/abs/1412.7412 , https://hal.archives-ouvertes.fr/hal-01098554

[34] A. AL GERBI, B. JOURDAIN, E. CLÉMENT.*Ninomiya-Victoir scheme : Multilevel Monte-Carlo estimators and discretization of the involved Ordinary Differential Equations*, in "ESAIM: Proceedings and Surveys", November 2017, vol. 59, p. 1-14, https://arxiv.org/abs/1612.07017 , https://hal.archives-ouvertes.fr/hal-01421337

[35] A. ALFONSI, P. BLANC.*Extension and calibration of a Hawkes-based optimal execution model*, in "Market microstructure and liquidity", August 2016, https://arxiv.org/abs/1506.08740 [*DOI :* 10.1142/S2382626616500052], https://hal-enpc.archives-ouvertes.fr/hal-01169686

[36] A. ALFONSI, J. CORBETTA, B. JOURDAIN.*Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems*, September 2017, preprint, https://hal.archives-ouvertes.fr/hal-01589581

[37] A. ALFONSI, M. HAYASHI, A. KOHATSU-HIGA.*Parametrix Methods for One-Dimensional Reflected SDEs*, in "Modern Problems of Stochastic Analysis and StatisticsSelected Contributions In Honor of Valentin Konakov", Springer, November 2017, vol. Springer Proceedings in Mathematics & Statistics, n$^{o}$ 208, https://hal-enpc.archives-ouvertes.fr/hal-01670011

[38] A. ALFONSI, A. KEBAIER, C. REY.*Maximum Likelihood Estimation for Wishart processes*, in "Stochastic Processes and their Applications", November 2016, https://arxiv.org/abs/1508.03323 [*DOI :* 10.1016/J.SPA.2016.04.026], https://hal-enpc.archives-ouvertes.fr/hal-01184310

[39] A. ALFONSI, A. SCHIED, F. KLÖCK.*Multivariate transient price impact and matrix-valued positive definite functions*, in "Mathematics of Operations Research", March 2016, https://arxiv.org/abs/1310.4471 [*DOI :* 10.1287/MOOR.2015.0761], https://hal-enpc.archives-ouvertes.fr/hal-00919895

[40] H. AMINI, R. CONT, A. MINCA.*Resilience to Contagion in Financial Networks*, in "Mathematical Finance", 2013

[41] H. AMINI, A. MINCA, A. SULEM.*Optimal equity infusions in interbank networks*, in "Journal of Financial Stability", August 2017, vol. 31, p. 1 - 17 [*DOI :* 10.1016/J.JFS.2017.05.008], https://hal.inria.fr/hal-01614759

[42] A. ARNOLD, E. CARLEN, Q. JU.*Large-time behavior of non-symmetric Fokker-Planck type equations*, in "Communications on Stochastic Analysis", 2008, vol. 2, n$^{o}$ 1, p. 153-175

[43] R. ASSARAF, B. JOURDAIN, T. LELIÈVRE, R. ROUX.*Computation of sensitivities for the invariant measure of a parameter dependent diffusion*, in "Stochastics and Partial Differential Equations: Analysis and Computations", October 2017, p. 1-59, https://arxiv.org/abs/1509.01348 [*DOI :* 10.1007/s40072-017-0105-6], https://hal.archives-ouvertes.fr/hal-01192862

[44] V. BALLY, L. CARAMELLINO.*Asymptotic development for the CLT in total variation distance*, in "Bernoulli", 2016, vol. 22, p. 2442-2485, https://hal-upec-upem.archives-ouvertes.fr/hal-01104866

[45] V. BALLY, L. CARAMELLINO.*Regularity of Wiener functionals under a Hörmander type condition of order one*, in "Annals of Probability", 2017, vol. 45, n$^{o}$ 3, p. 1488-1511, https://hal-upec-upem.archives-ouvertes.fr/hal-01413556

[46] V. BALLY, L. CARAMELLINO, P. PIGATO.*Diffusions under a local strong Hörmander condition. Part I: density estimates*, December 2016, preprint, https://hal-upec-upem.archives-ouvertes.fr/hal-01413546

[47] V. BALLY, L. CARAMELLINO, P. PIGATO.*Diffusions under a local strong Hörmander condition. Part II: tube estimates*, July 2016, preprint, https://hal.archives-ouvertes.fr/hal-01407420

[48] V. BALLY, L. CARAMELLINO, G. POLY.*Convergence in distribution norms in the CLT for non identical distributed random variables*, January 2017, preprint, https://hal-upec-upem.archives-ouvertes.fr/hal-01413548

[49] V. BALLY, L. CARAMELLINO, G. POLY.*Non universality for the variance of the number of real roots of random trigonometric polynomials*, 2017, preprint, https://hal.archives-ouvertes.fr/hal-01634848

[50] V. BALLY, A. KOHATSU-HIGA.*A probabilistic interpretation of the parametrix method*, in "Annals of Applied Probability", 2015, vol. 25, p. 3095-3138, https://hal.archives-ouvertes.fr/hal-00926479

[51] V. BALLY, V. RABIET.*Asymptotic behavior for multi-scale PDMP's*, April 2015, preprint, https://hal.archives-ouvertes.fr/hal-01144107

[52] V. BALLY, C. REY.*Approximation of Markov semigroups in total variation distance*, in "Electronic Journal of Probability", 2016, vol. 21, n° 12, https://hal-upec-upem.archives-ouvertes.fr/hal-01110015

[53] M. BEIGLBÖCK, P.-H. LABORDÈRE, F. PENKNER.*Model-independent bounds for option prices - a mass transport approach*, in "Finance Stoch.", 2013, vol. 17, n° 3, p. 477-501

[54] R. CHEN, A. MINCA, A. SULEM.*Optimal connectivity for a large financial network*, in "ESAIM: Proceedings and Surveys", 2017, vol. 59, p. 43 - 55, Editors : B. Bouchard, E. Gobet and B. Jourdain, https://hal.inria.fr/hal-01618701

[55] Q. DAI, K. J. SINGLETON.*Specification Analysis of Affine Term Structure Models*, in "The Journal of Finance", 2000, vol. 55, n° 5, p. 1943–1978, http://dx.doi.org/10.1111/0022-1082.00278

[56] D. DUFFIE, D. FILIPOVIĆ, W. SCHACHERMAYER.*Affine processes and applications in finance*, in "Ann. Appl. Probab.", 2003, vol. 13, n° 3, p. 984–1053, http://dx.doi.org/10.1214/aoap/1060202833

[57] R. DUMITRESCU, C. LABART.*Numerical approximation of doubly reflected BSDEs with jumps and RCLL obstacles*, in "Journal of Mathematical Analysis and applications", October 2016, vol. 442, n° 1, p. 206-243, https://arxiv.org/abs/1406.3612 , https://hal.archives-ouvertes.fr/hal-01006131

[58] R. DUMITRESCU, C. LABART.*Reflected scheme for doubly reflected BSDEs with jumps and RCLL obstacles*, in "Journal of Computational and Applied Mathematics", April 2016, vol. 296, p. 827-839, https://arxiv.org/abs/1502.02888 , https://hal.archives-ouvertes.fr/hal-01114996

[59] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM.*Optimal Stopping for Dynamic Risk Measures with Jumps and Obstacle Problems*, in "Journal of Optimization Theory and Applications", 2015, vol. 167, n° 1, 23 [*DOI :* 10.1007/s10957-014-0635-2], https://hal.inria.fr/hal-01096501

[60] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM.*Generalized Dynkin games and doubly reflected BSDEs with jumps*, in "Electronic Journal of Probability", 2016 [*DOI :* 10.1214/16-EJP4568], https://hal.inria.fr/hal-01388022

[61] R. DUMITRESCU, M.-C. QUENEZ, A. SULEM. *Mixed generalized Dynkin game and stochastic control in a Markovian framework*, in "Stochastics", 2016, vol. 89, n° 1, 30, https://hal.inria.fr/hal-01417203

[62] N. EL KAROUI, V. LACOSTE.*Multifactor models of the term structure of interest rates*, 1992, Preprint University of Paris 6

[63] A. FIGALLI.*Existence and uniqueness for martingale solutions of SDEs with rough or degenerate coefficients*, in "Journal of Functional Analysis", 2008, vol. 254, p. 109–153

[64] C. FONTANA, B. ØKSENDAL, A. SULEM.*Market viability and martingale measures under partial information*, in "Methodology and Computing in Applied Probability", 2015, vol. 17, 24 [*DOI :* 10.1007/s11009-014-9397-4], https://hal.inria.fr/hal-00789517

[65] J. FONTBONA, B. JOURDAIN.*A trajectorial interpretation of the dissipations of entropy and Fisher information for stochastic differential equations*, in "Annals of Probability", February 2016, vol. 44, n° 1, p. 131-170, https://arxiv.org/abs/1107.3300 , https://hal.archives-ouvertes.fr/hal-00608977

[66] G. FORT, B. JOURDAIN, E. KUHN, T. LELIÈVRE, G. STOLTZ.*Convergence of the Wang-Landau algorithm*, in "Mathematics of Computation", September 2015, vol. 84, n° 295 [*DOI :* 10.1090/S0025-5718-2015-02952-4], https://hal.inria.fr/hal-01238595

[67] G. FORT, B. JOURDAIN, E. KUHN, T. LELIÈVRE, G. STOLTZ.*Efficiency of the Wang-Landau Algorithm: A Simple Test Case*, in "Applied Mathematics Research eXpress", 2014, vol. 2014, p. 275-311 [*DOI :* 10.1093/AMRX/ABU003], https://hal.inria.fr/hal-00721886

[68] G. FORT, B. JOURDAIN, T. LELIÈVRE, G. STOLTZ.*Self-Healing Umbrella Sampling: Convergence and efficiency*, in "Statistics and Computing", January 2017, vol. 27, n° 1, p. 147–168, https://arxiv.org/abs/1410.2109 [*DOI :* 10.1007/s11222-015-9613-2], https://hal.archives-ouvertes.fr/hal-01073201

[69] N. FOURNIER, B. JOURDAIN.*Stochastic particle approximation of the Keller-Segel equation and two-dimensional generalization of Bessel processes*, in "The Annals of Applied Probability : an official journal of the institute of mathematical statistics", November 2017, vol. 27, n° 5, p. 2807-2861, https://arxiv.org/abs/1507.01087 , https://hal-enpc.archives-ouvertes.fr/hal-01171481

[70] J. GARNIER, G. PAPANICOLAOU, T. YANG.*Large deviations for a mean field model of systemic risk*, in "SIAM Journal on Financial Mathematics", 2013, vol. 41, n° 1, p. 151–184

[71] J. GUYON, P. HENRY-LABORDÈRE.*Being particular about calibration*, in "Risk", January 2012

[72] Y. HU, B. ØKSENDAL, A. SULEM.*Singular mean-field control games*, in "Stochastic Analysis and Applications", June 2017, vol. 35, n° 5, p. 823 - 851 [*DOI :* 10.1080/07362994.2017.1325745], https://hal.inria.fr/hal-01614747

[73] B. JOURDAIN, A. KEBAIER.*Non-asymptotic error bounds for The Multilevel Monte Carlo Euler method applied to SDEs with constant diffusion coefficient*, August 2017, preprint, https://hal.archives-ouvertes.fr/hal-01577874

[74] B. JOURDAIN, J. REYGNER.*The small noise limit of order-based diffusion processes*, in "Electronic Journal of Probability", March 2014, vol. 19, n° 29, p. 1-36, https://arxiv.org/abs/1307.0490 [*DOI :* 10.1214/EJP.v19-2906], https://hal-enpc.archives-ouvertes.fr/hal-00840185

[75] B. JOURDAIN, J. REYGNER.*Capital distribution and portfolio performance in the mean-field Atlas model*, in "Annals of Finance", May 2015, vol. 11, n$^{o}$ 2, p. 151-198, https://arxiv.org/abs/1312.5660 [*DOI :* 10.1007/s10436-014-0258-5], https://hal-enpc.archives-ouvertes.fr/hal-00921151

[76] B. JOURDAIN, J. REYGNER.*Optimal convergence rate of the multitype sticky particle approximation of one-dimensional diagonal hyperbolic systems with monotonic initial data*, in "Discrete and Continuous Dynamical Systems - Series A", September 2016, vol. 36, n$^{o}$ 9, p. 4963-4996, https://arxiv.org/abs/1507.01085 [*DOI :* 10.3934/DCDS.2016015], https://hal-enpc.archives-ouvertes.fr/hal-01171261

[77] B. JOURDAIN, A. ZHOU.*Fake Brownian motion and calibration of a Regime Switching Local Volatility model*, July 2016, preprint, https://hal.archives-ouvertes.fr/hal-01341212

[78] D. LÉPINGLE.*Euler scheme for reflected stochastic differential equations*, in "Math. Comput. Simulation", 1995, vol. 38, n$^{o}$ 1-3, p. 119–126, Probabilités numériques (Paris, 1992), http://dx.doi.org/10.1016/0378-4754(93)E0074-F

[79] A. MINCA.*Modélisation mathématique de la contagion de défaut; Mathematical modeling of financial contagion*, Université Pierre et Marie Curie, Paris 6, September 5 2011

[80] A. MINCA, A. SULEM.*Optimal Control of Interbank Contagion Under Complete Information*, in "Statistics and Risk Modeling", 2014, vol. 31, n$^{o}$ 1, p. 1001-1026 [*DOI :* 10.1524/STRM.2014.5005], https://hal.inria.fr/hal-00916695

[81] S. NINOMIYA, N. VICTOIR.*Weak approximation of stochastic differential equations and application to derivative pricing*, in "Applied Mathematical Finance", 2008, vol. 15, p. 107-121

[82] M.-C. QUENEZ, A. SULEM.*Reflected BSDEs and robust optimal stopping for dynamic risk measures with jumps*, in "Stochastic Processes and their Applications", September 2014, vol. 124, n$^{o}$ 9, 23, https://arxiv.org/abs/1212.6744 , https://hal.inria.fr/hal-00773708

[83] B. ØKSENDAL, A. SULEM.*Forward–Backward Stochastic Differential Games and Stochastic Control under Model Uncertainty*, in "Journal of Optimization Theory and Applications", April 2014, vol. 161, n$^{o}$ 1, p. 22 - 55 [*DOI :* 10.1007/s10957-012-0166-7], https://hal.inria.fr/hal-01681150

[84] B. ØKSENDAL, A. SULEM.*Dynamic Robust Duality in Utility Maximization*, in "Applied Mathematics and Optimization", 2016, p. 1-31, https://hal.inria.fr/hal-01406663

[85] B. ØKSENDAL, A. SULEM.*Optimal control of predictive mean-field equations and applications to finance*, in "Springer Proceedings in Mathematics & Statistics", Stochastic of Environmental and Financial Economics, Springer Verlag, 2016, vol. 138, p. 301–320 [*DOI :* 10.1007/978-3-319-23425-0], https://hal.inria.fr/hal-01406649

# Project-Team MIMOVE

## Middleware on the Move

RESEARCH CENTER
**Paris**

THEME
**Distributed Systems and middleware**

# Table of contents

# Project-Team MIMOVE

*Creation of the Team: 2014 July 01, updated into Project-Team: 2018 February 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
A1.3. - Distributed Systems
A1.4. - Ubiquitous Systems
A1.5. - Complex systems
A1.5.1. - Systems of systems
A1.5.2. - Communicating systems
A2.5. - Software engineering
A2.6.2. - Middleware
A3.1.7. - Open data
A3.1.8. - Big data (production, storage, transfer)
A3.3. - Data and knowledge analysis
A3.5. - Social networks

**Other Research Topics and Application Domains:**

B6.3. - Network functions
B6.4. - Internet of things
B6.5. - Information systems
B8.2. - Connected city
B8.5.1. - Participative democracy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Nikolaos Georgantas [Team leader, Inria, Researcher, HDR]
Vassilis Christophides [Inria, Advanced Research Position]
Renata Cruz Teixeira [Inria, & Stanford University from Sept 2018, Senior Researcher, HDR]
Valérie Issarny [Inria, & Inria@SiliconValley until Aug 2018, Senior Researcher, HDR]

**External Collaborators**

Françoise Sailhan [Associate Professor, CNAM]
Rafael Angarita Arocha [Associate Professor, ISEP]
Francesco Bronzino [Nokia Bell Labs, from Nov 2018]
Zied Ben Houidi [Nokia Bell Labs]
Diego Neves Da Hora [Google Brazil]

**Technical Staff**

Rachit Agarwal [Inria]
Ehsan Ahvar [Inria, from Jul 2018]
Shohreh Ahvar [Inria, from Jul 2018]
Giulio Grassi [Inria, from Apr 2018 until Sep 2018]
Bruno Lefèvre [Inria, from Apr 2018]
Patient Ntumba [Inria, until Jul 2018]
Pierre-Guillaume Raverdy [Inria & Inria SED, until Mar 2018]

**PhD Students**

Fethi Dilmi [Inria]
Yifan Du [Inria]
Grigorios Piperagkas [Inria, from Oct 2018]
Sarah Wassermann [Inria]
Patient Ntumba [Inria, from Aug 2018]
Israel Marquez Salinas [Université Pierre et Marie Curie, from Oct 2018]

**Post-Doctoral Fellows**

Sara Ayoubi [Inria]
Georgios Bouloukakis [Inria, & University of California, Irvine from Feb 2018]
Francesco Bronzino [Inria, until Oct 2018]
Giulio Grassi [Inria, from Oct 2018]
Bruno Lefèvre [Inria, until Mar 2018]

**Visiting Scientist**

Mark Crovella [Boston University, from Aug 2018]

**Administrative Assistant**

Nathalie Gaudechoux [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

Given the prevalence of global networking and computing infrastructures (such as the Internet and the Cloud), mobile networking environments, powerful hand-held user devices, and physical-world sensing and actuation devices, the possibilities of new mobile distributed systems have reached unprecedented levels. Such systems are dynamically composed of networked resources in the environment, which may span from the immediate neighborhood of the users – as advocated by pervasive computing – up to the entire globe – as envisioned by the Future Internet and one of its major constituents, the Internet of Things. Hence, we can now talk about truly ubiquitous computing.

The resulting ubiquitous systems have a number of unique – individually or in their combination – features, such as dynamicity due to volatile resources and user mobility, heterogeneity due to constituent resources developed and run independently, and context-dependence due to the highly changing characteristics of the execution environment, whether technical, physical or social. The latter two aspects are particularly manifested through the physical but also social sensing and actuation capabilities of mobile devices and their users. More specifically, leveraging the massive adoption of smart phones and other user-controlled mobile devices, besides physical sensing – where a device's sensor passively reports the sensed phenomena – *social sensing*/*crowd sensing* comes into play, where the user is aware of and indeed aids in the sensing of the environment.

Mobile systems with the above specifics further push certain problems related to the Internet and user experience to their extreme: (i) Technology is too complex. Most Internet users are not tech-savvy and hence cannot fix performance problems and anomalous network behavior by themselves. The complexity of most Internet applications makes it hard even for networking experts to fully diagnose and fix problems. Users can't even know whether they are getting the Internet performance that they are paying their providers for. (ii) There is too much content. The proliferation of user-generated content (produced anywhere with mobile devices and immediately published in social media) along with the vast amount of information produced by traditional media (e.g., newspapers, television, radio) poses new challenges in achieving an effective, near real-time information awareness and personalization. For instance, users need novel filtering and recommendation tools for helping them to decide which articles to read or which movie to watch.

This challenging context raises key research questions:

- How to deal with heterogeneity and dynamicity, which create runtime uncertainty, when developing and running mobile systems in the open and constantly evolving Internet and IoT environment?
- How to enable automated diagnosis and optimization of networks and systems in the Internet and IoT environment for improving the QoE of their users?
- How to raise human centric crowd-sensing to a reliable means of sensing world phenomena?
- How to deal with combination, analysis and privacy aspects of Web/social media and IoT crowd-sensing data streams?

# 3. Research Program

## 3.1. Introduction

The research questions identified above call for radically new ways in conceiving, developing and running mobile distributed systems. In response to this challenge, MiMove's research aims at enabling next-generation mobile distributed systems that are the focus of the following research topics.

## 3.2. Emergent mobile distributed systems

Uncertainty in the execution environment calls for designing mobile distributed systems that are able to run in a beforehand unknown, ever-changing context. Nevertheless, the complexity of such change cannot be tackled at system design-time. Emergent mobile distributed systems are systems which, due to their automated, dynamic, environment-dependent composition and execution, *emerge* in a possibly non-anticipated way and manifest *emergent properties*, i.e., both systems and their properties take their complete form only at runtime and may evolve afterwards. This contrasts with the typical software engineering process, where a system is finalized during its design phase. MiMove's research focuses on enabling the emergence of mobile distributed systems while assuring that their required properties are met. This objective builds upon pioneering research effort in the area of *emergent middleware* initiated by members of the team and collaborators [1], [3].

## 3.3. Large-scale mobile sensing and actuation

The extremely large scale and dynamicity expected in future mobile sensing and actuation systems lead to the clear need for algorithms and protocols for addressing the resulting challenges. More specifically, since connected devices will have the capability to sense physical phenomena, perform computations to arrive at decisions based on the sensed data, and drive actuation to change the environment, enabling proper coordination among them will be key to unlocking their true potential. Although similar challenges have been addressed in the domain of networked sensing, including by members of the team [7], the specific challenges arising from the *extremely large scale* of mobile devices – a great number of which will be attached to people, with uncontrolled mobility behavior – are expected to require a significant rethink in this domain. MiMove's research investigates techniques for efficient coordination of future mobile sensing and actuation systems with a special focus on their dependability.

## 3.4. Mobile social crowd-sensing

While mobile social sensing opens up the ability of sensing phenomena that may be costly or impossible to sense using embedded sensors (e.g., subjective crowdedness causing discomfort or joyfulness, as in a bus or in a concert) and leading to a feeling of being more socially involved for the citizens, there are unique consequent challenges. Specifically, MiMove's research focuses on the problems involved in the combination of the physically sensed data, which are quantitative and objective, with the mostly qualitative and subjective data arising from social sensing. Enabling the latter calls for introducing mechanisms for incentivising user participation and ensuring the privacy of user data, as well as running empirical studies for understanding the complex social behaviors involved. These objectives build upon previous research work by members of the team on mobile social ecosystems and privacy, as well as a number of efforts and collaborations in the domain of smart cities and transport that have resulted in novel mobile applications enabling empirical studies of social sensing systems.

## 3.5. Active and passive probing methods

We are developing methods that actively introduce probes in the network to discover properties of the connected devices and network segments. We are focusing in particular on methods to discover properties of home networks (connected devices and their types) and to distinguish if performance bottlenecks lie within the home network versus in the different network segments outside (e.g., Internet access provider, interconnects, or content provider). Our goal is to develop adaptive methods that can leverage the collaboration of the set of available devices (including end-user devices and the home router, depending on which devices are running the measurement software).

We are also developing passive methods that simply observe network traffic to infer the performance of networked applications and the location of performance bottlenecks, as well as to extract patterns of web content consumption. We are working on techniques to collect network traffic both at user's end-devices and at home routers. We also have access to network traffic traces collected on a campus network and on a large European broadband access provider.

## 3.6. Inferring user online experience

We are developing hybrid measurement methods that combine passive network measurement techniques to infer application performance with techniques from HCI to measure user perception. We will later use the resulting datasets to build models of user perception of network performance based only on data that we can obtain automatically from the user device or from user's traffic observed in the network.

## 3.7. Real time data analytics

The challenge of deriving insights from the Internet of Things (IoT) has been recognized as one of the most exciting and key opportunities for both academia and industry. The time value of data is crucial for many IoT-based systems requiring *real-time* (or near real-time) *control* and *automation*. Such systems typically collect data continuously produced by "things" (i.e., devices), and analyze them in (sub-) seconds in order to act promptly, e.g., for detecting security breaches of digital systems, for spotting malfunctions of physical assets, for recommending goods and services based on the proximity of potential clients, etc. Hence, they require to both *ingest* and *analyze in real-time* data arriving with different velocity from various IoT data streams.

Existing incremental (online or streaming) techniques for descriptive statistics (e.g., frequency distributions, frequent patterns, etc.) or predictive statistics (e.g., classification, regression) usually assume a good enough quality dataset for mining patterns or training models. However, IoT raw data produced in the wild by sensors embedded in the environment or wearable by users are prone to errors and noise. Effective and efficient algorithms are needed for *detecting* and *repairing data impurities* (for controlling data quality) as well as *understanding data dynamics* (for defining alerts) in real-time, for collections of IoT data streams that might be geographically distributed. Moreover, supervised deep learning and data analytics techniques are challenged

by the presence of sparse ground truth data in real IoT applications. Lightweight and adaptive semi-supervised or unsupervised techniques are needed to power real-time anomaly and novelty detection in IoT data streams. The effectiveness of these techniques should be able to reach a useful level through training on a relatively small amount of (preferably unlabeled) data while they can cope distributional characteristics of data evolving over time.

# 4. Application Domains

## 4.1. Mobile urban systems for smarter cities

With the massive scale adoption of mobile devices and further expected significant growth in relation with the Internet of Things, mobile computing is impacting most – if not all – the ICT application domains. However, given the importance of conducting empirical studies to assess and nurture our research, we focus on one application area that is the one of "*smart cities*". The smart city vision anticipates that the whole urban space, including buildings, power lines, gas lines, roadways, transport networks, and cell phones, can all be wired together and monitored. Detailed information about the functioning of the city then becomes available to both city dwellers and businesses, thus enabling better understanding and consequently management of the city's infrastructure and resources. This raises the prospect that cities will become more sustainable environments, ultimately enhancing the citizens' well being. There is the further promise of enabling radically new ways of living in, regulating, operating and managing cities, through the increasing active involvement of citizens by ways of crowd-sourcing/sensing and social networking.

Still, the vision of what smart cities should be about is evolving at a fast pace in close concert with the latest technology trends. It is notably worth highlighting how mobile and social network use has reignited citizen engagement, thereby opening new perspectives for smart cities beyond data analytics that have been initially one of the core foci for smart cities technologies. Similarly, open data programs foster the engagement of citizens in the city operation and overall contribute to make our cities more sustainable. The unprecedented democratization of urban data fueled by open data channels, social networks and crowd sourcing enables not only the monitoring of the activities of the city but also the assessment of their nuisances based on their impact on the citizens, thereby prompting social and political actions. However, the comprehensive integration of urban data sources for the sake of sustainability remains largely unexplored. This is an application domain that we focus on, further leveraging our research on emergent mobile distributed systems, large-scale mobile sensing & actuation, and mobile social crowd-sensing.

In particular, we concentrate on the following specialized applications, which we have investigated in close collaboration with other researchers as part of the dedicated Inria Project Lab *CityLab@Inria*:

- **Democratization of urban data for healthy cities.** We integrate the various urban data sources, especially by way of crowd-Xing, to better understand city nuisances. This goes from raw pollution sensing (e.g., sensing noise) to the sensing of its impact on citizens (e.g., how people react to urban noise and how this affects their health).
- **Social applications.** Mobile applications are being considered by sociologists as a major vehicle to actively involve citizens and thereby prompt them to become activists. We study such a vehicle from the ICT perspective and in particular elicit relevant middleware solutions to ease the development of such "*civic apps*".

More specifically, MiMove led CityLab@Inria [0] from Jan 2014 to Nov 2018. CityLab focused on the study of ICT solutions promoting social sustainability in smart cities, and involved the following Inria project-teams in addition to MiMove: CLIME/ANGE, DICE, FUN, MYRIADS, SMIS/PETRUS, URBANET/AGORA. CityLab further involved strong collaboration with California universities affiliated with CITRIS (Center for Information Technology Research in the Interest of Society) and especially UC Berkeley, in relation with the *Inria@SiliconValley* program.

[0] http://citylab.inria.fr

## 4.2. Home network diagnosis

With the availability of cheap broadband connectivity, Internet access from the home has become a ubiquity. Modern households host a multitude of networked devices, ranging from personal devices such as laptops and smartphones to printers and media centers. These devices connect among themselves and to the Internet via a local-area network–a home network–that has become an important part of the "Internet experience". In fact, ample anecdotal evidence suggests that the home network can cause a wide array of connectivity impediments, but their nature, prevalence, and significance remain largely unstudied.

Our long-term goal is to assist users with concrete indicators of the quality of their Internet access, causes of potential problems and–ideally–ways to fix them. We intend to develop a set of easy-to-use home network monitoring and diagnosis tools. The development of home network monitoring and diagnosis tools brings a number of challenges. First, home networks are heterogeneous. The set of devices, configurations, and applications in home networks vary significantly from one home to another. We must develop sophisticated techniques that can learn and adapt to any home network as well as to the level of expertise of the user. Second, Internet application and services are also heterogeneous with very diverse network requirements. We must develop methods that can infer application quality solely from the observation of (often encrypted) application network traffic. There are numerous ways in which applications can fail or experience poor performance in home networks. Often there are a number of explanations for a given symptom. We must devise techniques that can identify the most likely cause(s) for a given problem from a set of possible causes. Finally, even if we can identify the cause of the problem, we must then be able to identify a solution. It is important that the output of the diagnosis tools we build is "actionable". Users should understand the output and know what to do.

We are working with Princeton University (associate team HOMENET) to deploy monitoring infrastructure within users' homes. Our goal is to develop a mostly passive measurement system to monitor the performance of user applications, which we call NetMicroscope. We are developing NetMicroscope to run in a box acting as home gateway. Our current deployments use Raspberry Pi and Odroid boxes. We have these boxes deployed in 50 homes in the US and 10 in France. The US deployment is run and financed by the Wall Street Journal. We are collaborating with them to understand the relationship between Internet access speed and video quality. We have been discussing with Internet regulators (in particular, FCC, ACERP, and BEREC) as well as residential access ISP in how NetMicroscope can help overcome the shortcomings of existing Internet quality monitoring systems.

## 4.3. Mobile Internet quality of experience

Mobile Internet usage has boomed with the advent of ever smarter handheld devices and the spread of fast wireless access. People rely on mobile Internet for everyday tasks such as banking, shopping, or entertainment. The importance of mobile Internet in our lives raises people's expectations. Ensuring good Internet user experience (or Quality of Experience—QoE) is challenging, due to the heavily distributed nature of Internet services. For mobile applications, this goal is even more challenging as access connectivity is less predictable due to user mobility, and the form factor of mobile devices limits the presentation of content. For these reasons, the ability to monitor QoE metrics of mobile applications is essential to determine when the perceived application quality degrades and what causes this degradation in the chain of delivery. Our goal is to improve QoE of mobile applications.

To achieve this goal, we are working on three main scientific objectives. First, we are working on novel methods to monitor mobile QoE. Within the IPL BetterNet we are developing the HostView for Android tool that runs directly on mobile devices to monitor network and system performance together with the user perception of performance. Second, we plan to develop models to predict QoE of mobile applications. We will leverage the datasets collected with HostView for Android to build data-driven models. Finally, our goal is to develop methods to optimize QoE for mobile users. We are currently developing optimization methods for interactive video applications. We envision users walking or driving by road-side WiFi access points (APs) with full 3G/LTE coverage and patchy WiFi coverage (i.e., community Wifi or Wifi APs on Lampposts). To

achieve this goal, we plan to leverage multi-path and cross-layer optimizations. We have started conducting experiments in the Paris subway and walking around Inria to measure the quality of FreeWiFi as well as LTE connectivity. We are experimenting with existing multipath protocols (MP-TCP and MP-DASH). We are also analyzing connectivity in datasets from the MONROE project (which measure LTE in Europe) and CarFi (which measures WiFi quality from APs deployed in cars).

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

- Renata Teixeira was named ACM Distinguished member for outstanding scientific contributions to computing in 2018.
- Our paper "Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics" received the Best Dataset Award at the Passive and Active Measurement Conference 2018.

BEST PAPERS AWARDS :
[27]
R. GOMES, G. BOULOUKAKIS, F. COSTA, N. GEORGANTAS, R. DA ROCHA.*QoS-Aware Resource Allocation for Mobile IoT Pub/Sub Systems*, in "2018 International Conference on Internet of Things (ICIOT)", Seattle, United States, June 2018, https://hal.inria.fr/hal-01797933

# 6. New Software and Platforms

## 6.1. SocialBus

*Universal Social Network Bus*
KEYWORDS: Middleware - Interoperability - Social networks - Software Oriented Service (SOA)
FUNCTIONAL DESCRIPTION: Online social network services (OSNSs) have become an integral part of our daily lives. At the same time, the aggressive market competition has led to the emergence of multiple competing siloed OSNSs that cannot interoperate. As a consequence, people face the burden of creating and managing multiple OSNS accounts and learning how to use them, to stay connected. The goal of the Universal Social Network Bus (USNB) is to relieve users from such a burden, letting them use their favorite applications to communicate.

- Authors: Rafael Angarita Arocha, Nikolaos Georgantas and Valérie Issarny
- Contact: Valérie Issarny
- URL: https://gitlab.inria.fr/usnb/universal-social-network-bus

## 6.2. WeBrowse

KEYWORDS: Web Usage Mining - Content analysis - Recommendation systems
FUNCTIONAL DESCRIPTION: The amount of information available on the web today, and the fast rate with which new information appears, overwhelm most users. The goal of our research is to assist Web users in discovering content. One of the most powerful means today to help people discover new web content is sharing between members of online communities. In the case of communities of a place (e.g., people who live, study, or work together) people share common interests, but often fail to actively share content. To address this problem, we have developed WeBrowse, a passive crowdsourced content discovery system for communities of a place.

WeBrowse leverages the passive observation of web-clicks (i.e., the URLs users intentionally visit) as an indication of users' interest in a piece of content. Intuitively, the more users click on a URL, the higher the interest in the content on the corresponding page. Our approach is then to leverage the collective clicks in a community to automatically discover relevant content to promote to users of the community.

To implement passive crowdsourcing, one must be in a position to observe the aggregated web-clicks of the community. Luckily, in many communities of a place, users will connect to the Internet from the same network, such as, e.g., the campus/enterprise network or the network of a residential Internet Service Provider (ISP) in a neighborhood. WeBrowse (i) observes web packets flowing through a network link, (ii) passively extracts HTTP logs (i.e., streams recording the headers of HTTP requests), and (iii) detects and decides on-the-fly the set of URLs to show to users.

- Contact: Renata Cruz Teixeira
- URL: https://team.inria.fr/muse/webrowse-info-page/

## 6.3. EEE

*Experiment Execution Engine*
KEYWORD: Iot
FUNCTIONAL DESCRIPTION: Experiment Execution Engine (EEE) eases the development of IoT applications that perform analysis of recent or continuously increasing volumes of IoT data from various data stores. To this end, EEE provides APIs for scheduling queries on federated large-scale semantically-enabled IoT data stores. Queries are expressed in the FIESTA-IoT (http://fiesta-iot.eu) Experiment Description Specification (FEDSpec), which acts as a Domain Specific Language (DSL). EEE (in combination with Experiment Management Console - EMC) further provides features such as: (a) manage scheduled experiments, (b) (un)subscribe to already existing FEDSpec objects, (c) execute an on-demand query outside a described schedule, (d) monitor execution logs, (e) interact with third-party services such as Analytics and Result Storage, (e) handle dynamic attributes, and (f) API sandbox. EEE is integrated within the FIESTA-IoT Platform. It can be customised depending on needs.

- Authors: Rachit Agarwal and Nikolaos Georgantas
- Contact: Nikolaos Georgantas
- URL: https://github.com/fiesta-iot/ExperimentExecutionEngine

## 6.4. EMC

*Experiment Management Console*
KEYWORD: Iot
FUNCTIONAL DESCRIPTION: Experiment Management Console (EMC) is a simple easy-to-use user interface that enables experimenters to control their experiments/queries that execute using the Experiment Execution Engine (EEE). EMC provides features such as managing experiment execution, (un)subscribing an existing experiment, and monitoring execution logs.

- Authors: Nikolaos Georgantas and Rachit Agarwal
- Contact: Nikolaos Georgantas
- URL: https://github.com/fiesta-iot/ExperimentManagementConsole

## 6.5. VSB

*eVolution Service Bus*
KEYWORDS: Service and Thing choreographies - Middleware protocol interoperability - Enterprise service bus

FUNCTIONAL DESCRIPTION: VSB is a development and runtime environment dedicated to complex distributed applications of the Future Internet. Such applications are open, dynamic choreographies of extremely heterogeneous services and Things, including lightweight embedded systems (e.g., sensors, actuators and networks of them), mobile systems (e.g., smartphone applications), and resource-rich IT systems (e.g., systems hosted on enterprise servers and Cloud infrastructures). VSB's objective is to seamlessly interconnect, inside choreographies, services and Things that employ heterogeneous interaction protocols at the middleware level, e.g., SOAP Web services, REST Web services, Things using CoAP. This is based on runtime conversions between such protocols, with respect to their primitives and data type systems, while properly mapping between their semantics. This also includes mapping between the public interfaces of services/Things, regarding their operations and data, from the viewpoint of the middleware: the latter means that operations and data are converted based on their middleware-level semantics, while their business semantics remains transparent to the conversion. VSB follows the well-known Enterprise Service Bus (ESB) paradigm. We propose a generic interface description, which we call GIDL, for application components that employ VSB. Based on GIDL, we enable automated synthesis of binding components for connecting heterogeneous services and Things onto VSB.

- Participants: Georgios Bouloukakis, Nikolaos Georgantas and Patient Ntumba
- Contact: Nikolaos Georgantas
- URL: https://gitlab.ow2.org/chorevolution/evolution-service-bus

## 6.6. Service traceroute

KEYWORDS: Network monitoring - Network diagnosis
FUNCTIONAL DESCRIPTION: Traceroute is often used to help diagnose when users experience issues with Internet applications or services. Unfortunately, probes issued by classic traceroute tools differ from application traffic and hence can be treated differently by middleboxes within the network. We propose a new traceroute tool, called Service traceroute. Service traceroute leverages the idea from paratrace, which passively listens to application traffic to then issue traceroute probes that pretend to be part of the application flow. We extend this idea to work for modern Internet services with support for identifying the flows to probe automatically, for tracing of multiple concurrent flows, and for UDP flows. We implement command-line and library versions of Service traceroute, which we release as open source.

- Partner: Princeton University
- Contact: Renata Cruz Teixeira
- URL: https://github.com/wontoniii/service-traceroute

## 6.7. TA

*TA - Traffic Analysis*
KEYWORDS: Quality of Experience - Network monitoring - Video analysis
FUNCTIONAL DESCRIPTION: System running at the home getaway that analyzes traffic generated by DASH on-demand and live video streams. The system tracks traffic patterns to infer key video QoE metrics such as average bitrate and re-buffering events. Moreover, the system exploits novel algorithms that use probing techniques, i.e. lightweight pings and traceroutes, to detect possible congestion location.

- Participants: Francesco Bronzino and Renata Cruz Teixeira
- Contact: Francesco Bronzino

## 6.8. HostView Mobile

KEYWORDS: Quality of Experience - Network monitoring
FUNCTIONAL DESCRIPTION: HostView for mobile runs on Android devices to monitor user system and network performance together with user feedback on Internet experience.

- Contact: Giulio Grassi

# 7. New Results

## 7.1. Ontology categorization for IoT semantics

**Participants:** Rachit Agarwal, Nikolaos Georgantas, Valérie Issarny.

IoT systems are now being deployed worldwide to sense phenomena of interest. The existing IoT systems are often independent which limits the use of sensor data to only one application. Semantic solutions have been proposed to support reuse of sensor data across IoT systems and applications. This allows integration of IoT systems for increased productivity by solving challenges associated with their interoperability and heterogeneity. Several ontologies have been proposed to handle different aspects of sensor data collection in IoT systems, ranging from sensor discovery to applying reasoning on collected sensor data for drawing inferences. In this work, we study and categorise the existing ontologies based on the fundamental ontological concepts (e.g., sensors, context, location, and more) required for annotating different aspects of data collection and data access in an IoT application. We identify these fundamental concepts by answering the 4Ws (What, When, Who, Where) and 1H (How) identified using the 4W1H methodology.

## 7.2. Massively-Parallel Feature Selection for Big Data

**Participant:** Vassilis Christophides.

We present the Parallel, Forward-Backward with Pruning (PFBP) algorithm for feature selection (FS) in Big Data settings (high dimensionality and/or sample size). To tackle the challenges of Big Data FS, PFBP partitions the data matrix both in terms of rows (samples, training examples) as well as columns (features). By employing the concepts of p-values of conditional independence tests and meta-analysis techniques, PFBP manages to rely only on computations local to a partition while minimizing communication costs. Then, it employs powerful and safe (asymptotically sound) heuristics to make early, approximate decisions, such as Early Dropping of features from consideration in subsequent iterations, Early Stopping of consideration of features within the same iteration, or Early Return of the winner in each iteration. PFBP provides asymptotic guarantees of optimality for data distributions faithfully representable by a causal network (Bayesian network or maximal ancestral graph). Our empirical analysis confirms a superlinear speedup of the algorithm with increasing sample size, linear scalability with respect to the number of features and processing cores, while dominating other competitive algorithms in its class.

## 7.3. Universal Social Network Bus

**Participants:** Ehsan Ahvar, Shohreh Ahvar, Rafael Angarita, Nikolaos Georgantas, Valérie Issarny, Bruno Lefèvre.

Online social network services (OSNSs) are changing the fabric of our society, impacting almost every aspect of it. Over the last decades, the aggressive market rivalry has led to the emergence of multiple competing, "closed" OSNSs. As a result, users are trapped in the walled gardens of their OSNS, encountering restrictions about what they can do with their personal data, the people they can interact with and the information they get access to. As an alternative to the platform lock-in, "open" OSNSs promote the adoption of open, standardized APIs. However, users still massively adopt closed OSNSs to benefit from the services' advanced functionalities and/or follow their "friends", although the users' virtual social sphere is ultimately limited by the OSNSs they join. Our work aims at overcoming such a limitation by enabling users to meet and interact beyond the boundary of their OSNSs, including reaching out to "friends" of distinct closed OSNSs. We specifically introduce USNB -*Universal Social Network Bus*, which revisits the "service bus" paradigm that enables interoperability across computing systems, to address the requirements of "*social interoperability*". USNB features *synthetic profiles* and *personae* for interaction across the boundaries of –closed and open–, –profile- and non-profile-based– OSNSs through a *reference social interaction service*.

USNB enables users to reach out to their social peers independently of the communication service (and especially underlying platform) each one uses in the virtual world. The success and massive adoption of OSNSs -as magnified by the success of Facebook- shows that online social communication is an essential tool for people. This further paves the way for collective and collaborative actions at the Internet scale. However, existing online collaborative tools come along with their communication platform, which is either a proprietary solution or a third-party OSNS. We argue that USNB contributes to enabling participatory systems at a larger inclusive scale by overcoming the technical boundaries set by existing online communication platforms. In that direction, we investigate the customization of USNB for specific applications and more specifically: participatory systems and massive open online courses.

## 7.4. Middleware for Mobile Crowdsensing

**Participants:** Yifan Du, Valérie Issarny, Bruno Lefèvre, Françoise Sailhan.

Mobile Phone Sensing (MPS) offers a great opportunity toward the large scale monitoring of urban phenomena, such as the exposition of the population to environmental pollution. Indeed, mobile crowdsensing empowers ordinary citizens to contribute (whether pro-actively or passively) data sensed or generated from their mobile devices. It allows acquiring hyperlocal knowledge at scale, thanks to the proliferation of mobile devices and the ubiquity of wireless broadband connection. On-demand mobile crowdsensing is in particular a cost-effective service model for smart cities. Numerous sensor types embedded in today's smartphones contribute valuable quantitative observations about the urban environment (e.g., noise, temperature, atmospheric pressure, humidity, light, magnetism). The observations further come along with the related spatial and temporal data, which allows for the analysis of hyper-local environmental knowledge. However, mobile crowdsensing brings valuable knowledge only if a sufficiently large crowd contributes and if we overcome the relatively low accuracy of the gathered data. This is the focus of our research.

We have in particular studied how to reduce the gap between the need for the massive collection of relevant data, and the quantity and accuracy of the measurements that are actually gathered. We specifically carried out an iterative research process to tackle this challenge, which combines technological innovation and social design. We have been developing a number of social tools to study the motivations and usages of MPS-based smart city apps, with the Ambiciti app serving as our use case. Our study has been taking into account the cultural and societal contexts that the usages of Ambiciti could feed, spanning health, environment, education, and urban policies. We carried out an online survey together with interviews with users and local actors in Europe, i.e., France, Belgium, and Finland. The research results contribute to a better understanding of why and how people use mobile phone sensing applications; the results also inform how to best leverage mobile crowd-sensing in the development of smart cities and how it may serve addressing urban challenges related to, e.g., public health or urban planning.

The quality of the contributed measurements challenges the aggregation of relevant knowledge from crowdsensed observations. The measurements quality depends on the *accuracy* of the contributing sensors and the adequacy of the *sensing context*. Addressing the former relies on the sensor calibration for which we study both micro- and macro-level solutions. Addressing the latter requires a supporting inference mechanism, for which we introduce a *personalized hierarchical inference* of all the context elements that are relevant to the phenomenon that is monitored through crowdsensing, and under which the crowdsensor operates. This enables accounting for the specific behavior of the contributing end-user across time, as well as for all the features -and only those- that are relevant and locally available, while reducing the feedback required from the user for the personalization.

## 7.5. QoS-Aware Resource Allocation for Mobile IoT Pub/Sub Systems

**Participants:** Georgios Bouloukakis, Nikolaos Georgantas.

IoT applications are usually characterized by large-scale demand and the widespread use of mobile devices. Similarly, performing interaction among application and system components in a decoupled and elastic way, and enforcing Quality of Service (QoS) usually also become issues. Hence, paradigms such as pub/sub on top

of cloud resources represent a suitable strategy for application development. However, management of QoS-aware resource allocation for pub/sub systems remains challenging, especially when system peers connect in an intermittent way. In this work, we propose a new approach for resource allocation focusing on end-to-end performance in face of peers' disconnections. We evaluate and demonstrate the benefits of our approach using simulations. QoS enforcement was achieved in almost all scenarios, and we have shown that our approach can help reasoning about efficient resource allocation.

## 7.6. Queueing Network Modeling Patterns for Reliable & Unreliable Pub/Sub Protocols

**Participants:** Georgios Bouloukakis, Nikolaos Georgantas, Patient Ntumba, Valérie Issarny.

Mobile Internet of Things (IoT) applications are typically deployed on resource-constrained devices with intermittent network connectivity. To support the deployment of such applications, the Publish/Subscribe (pub/sub) interaction paradigm is often employed, as it decouples mobile peers in time and space. Furthermore, pub/sub middleware protocols and APIs consider the Things' hardware limitations and support the development of effective applications by providing Quality of Service (QoS) features. These features aim to enable developers to tune an application by switching different levels of response times and delivery success rates. However, the profusion of pub/sub middleware protocols coupled with intermittent network connectivity result in non-trivial application tuning. In this work, we model the performance of middleware protocols found in IoT, which are classified within the pub/sub interaction paradigm – both reliable and unreliable underlying network layers are considered. We model reliable and unreliable protocols, by considering QoS semantics for data validity, buffer capacities, as well as the intermittent availability of peers. To this end, we rely on queueing network models, which offer a simple modeling environment that can be used to represent IoT interactions by combining multiple queueing model types. Based on these models, we perform statistical analysis by varying the QoS semantics, demonstrating their significant effect on response times and on the rate of successful interactions. We showcase the application of our analysis in concrete scenarios relating to Traffic Information Management systems, that integrate both reliable and unreliable participants. The consequent PerfMP performance modeling pattern may be tailored for a variety of deployments, in order to control fine-grained QoS policies.

## 7.7. Lightweight, General Inference of Streaming Video Quality from Encrypted Traffic

**Participants:** Francesco Bronzino, Sara Ayoubi, Renata Teixeira, Sarah Wasserman.

Accurately monitoring application performance is becoming more important for Internet Service Providers (ISPs), as users increasingly expect their networks to consistently deliver acceptable application quality. At the same time, the rise of end-to-end encryption makes it difficult for network operators to determine video stream quality–including metrics such as startup delay, resolution, rebuffering, and resolution changes–directly from the traffic stream. This work develops general methods to infer streaming video quality metrics from encrypted traffic using lightweight features. Our evaluation shows that our models are not only as accurate as previous approaches, but they also generalize across multiple popular video services, including Netflix, YouTube, Amazon Instant Video, and Twitch. The ability of our models to rely on lightweight features points to promising future possibilities for implementing such models at a variety of network locations along the end-to-end network path, from the edge to the core.

## 7.8. Service traceroute: Tracing Paths of Application Flows

**Participants:** Ivan Morandi, Francesco Bronzino, Renata Teixeira.

Traceroute is often used to help diagnose when users experience issues with Internet applications or services. Unfortunately, probes issued by classic traceroute tools differ from application traffic and hence can be treated differently by middleboxes within the network. This work proposes a new traceroute tool, called Service traceroute. Service traceroute leverages the idea from paratrace, which passively listens to application traffic

to then issue traceroute probes that pretend to be part of the application flow. We extend this idea to work for modern Internet services with support for identifying the flows to probe automatically, for tracing of multiple concurrent flows, and for UDP flows. We implement command-line and library versions of Service traceroute, which we release as open source. This paper also presents an evaluation of Service traceroute when tracing paths traversed by Web downloads from the top-1000 Alexa websites and by video sessions from Twitch and Youtube. Our evaluation shows that Service traceroute has no negative effect on application flows. Our comparison with Paris traceroute shows that a typical traceroute tool that launches a new flow to the same destination discovers different paths than when embedding probes in the application flow in a significant fraction of experiments (from 40% to 50% of our experiments in PlanetLab Europe).

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Grants with Industry

"Application Performance Bottleneck Detection", Comcast Gift to R. Teixeira 2018.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

"BottleNet: Understanding and Diagnosing End-to-end Communication Bottlenecks of the Internet", project funded by the French research agency (ANR), from Feb 2016 to Sep 2020.

### *9.1.1. Inria Support*

*9.1.1.1. Inria IPL CityLab@Inria*

**Participants:** Valérie Issarny, Bruno Lefèvre.

- **Name:** CityLab@Inria – *Overcoming the Smart City Challenge – Toward Environmental and Social Sustainability*
- **Period:** [January 2014 – November 2018]
- **Inria teams:** CLIME/ANGE, DICE, FUN, MIMOVE, MYRIADS, SMIS/PETRUS, URBANET/AGORA
- **URL:** http://citylab.inria.fr

The Inria Project Lab (IPL) 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 experiments is a central concern of the Lab, so that we are able to confront proposed approaches to actual settings.

*9.1.1.2. Inria IPL BetterNet*

**Participants:** Renata Teixeira, Vassilis Christophides, Francesco Bronzino.

- **Name:** BetterNet – *An observatory to measure and improve Internet service access from user experience*
- **Period:** [2016 – 2019]
- **Inria teams:** Diana, Dionysos, Inria Chile, Madynes, MiMove, Spirals
- **URL:** https://project.inria.fr/betternet/

BetterNet aims at building and delivering a scientific and technical collaborative observatory to measure and improve the Internet service access as perceived by users. In this Inria Project Lab, we will propose new original user-centered measurement methods, which will associate social sciences to better understand Internet usage and the quality of services and networks. Our observatory can be defined as a vantage point, where:

1. tools, models and algorithms/heuristics will be provided to collect data,
2. acquired data will be analyzed, and shared appropriately with scientists, stakeholders and civil society,
3. and new value-added services will be proposed to end-users.

*9.1.1.3. Inria ADT MOSQUITO*

**Participants:** Renata Teixeira, Francesco Bronzino.

- **Name:** MOSQUITO – *A mobile platform to measure the quality of Internet connectivity*
- **Period:** [November 2016 – October 2018]
- **Partners:** Inria MiMove, Inria SPIRALS.

The ADT MOSQUITO is part of the Inria Project Lab (IPL) initiative BetterNet. This ADT project focuses on the design and the development of a measurement platform for the quality of mobile Internet access by federating the existing mobile platforms identified in the BetterNet IPL. Beyond the priceless value of such a measurement platform for the research community, this ADT also aims to publish live reports on the quality of mobile Internet access through the BetterNet initiative.

## 9.2. European Initiatives

### *9.2.1. FP7 & H2020 Projects*

*9.2.1.1. H2020 ICT FIESTA-IoT*

**Participants:** Valérie Issarny, Nikolaos Georgantas, Rachit Agarwal.

**Name:** FIESTA-IoT – *Federated Interoperable Semantic IoT/cloud Testbeds and Applications*

**URL:** http://fiesta-iot.eu

**Type:** Research & Innovation Action (ICT)

**Topic:** FIRE+ (Future Internet Research & Experimentation)

**Period:** [February 2015 - January 2018]

**Partners:** Fraunhofer FOKUS (Germany) [**coordinator**], INSIGHT @ National University of Galway (Ireland) [**co-coordinator**], University of Southampton IT Innovation Centre (UK), Inria MiMove, University of Surrey (UK), Unparallel Innovation Lda (Portugal), Easy Global Market (France), NEC Europe Ltd (UK), University of Cantabria (Spain), Com4innov (France), Athens Information Technology (Greece), SOCIEDAD PARA EL DESARROLLO REGIONAL DE CANTABRIA (Spain), Ayuntamiento de Santander (Spain), Korea Electronics Technology Institute (Korea).

Despite the proliferation of IoT and smart cities testbeds, there is still no easy way to conduct large scale experiments that leverage data and resources from multiple geographically and administratively distributed IoT platforms. Recent advances in IoT semantic interoperability provide a sound basis for implementing novel cloud-based infrastructures that could allow testbed-agnostic access to IoT data and resources. FIESTA will open new horizons in IoT experimentation at a global scale, based on the interconnection and interoperability of diverse IoT testbeds. FIESTA will produce a first-of-a-kind blueprint experimental infrastructure (tools, techniques and best practices) enabling testbed operators to interconnect their facilities in an interoperable way, while at the same time facilitating researchers in deploying integrated experiments, which seamlessly transcend the boundaries of multiple IoT platforms. FIESTA will be validated and evaluated based on the interconnection of four testbeds (in Spain, UK, France and Korea), as well as based on the execution of novel experiments in the areas of mobile crowd-sensing, IoT applications portability, and dynamic intelligent discovery of IoT resources. In order to achieve global outreach and maximum impact, FIESTA will integrate an additional testbed and experiments from Korea, while it will also collaborate with IoT experts from USA. The participation of a Korean partner (based its own funding) will maximize FIESTA's value for EC money. Moreover, the project will take advantage of open calls processes towards attracting third-parties that will engage in the integration of their platforms within FIESTA or in the conduction of added-value experiments. As part of its sustainability strategy, FIESTA will establish a global market confidence programme for IoT interoperability, which will enable innovative platform providers and solution integrators to ensure/certify the openness and interoperability of their developments.

## 9.3. International Initiatives

### *9.3.1. Inria International Labs*

**Inria@SiliconValley**

Associate Team involved in the International Lab:

#### *9.3.1.1. MINES*

Title: Adaptive Communication Middleware for Resilient Sensing & Actuation IN Emergency Response Scenarios

International Partner:

University of California, Irvine (United States) - Information and Computer Science - Nalini Venkatasubramanian

Start year: 2018

See also: http://mimove-apps.paris.inria.fr/mines/index.html

Emerging smart-city and smart-community efforts will require a massive deployment of connected entities (Things) to create focused smartspaces. Related applications will enhance citizen quality of life and public safety (e.g., providing safe evacuation routes in fires). However, supporting IoT deployments are heterogeneous and can be volatile and failure-prone as they are often built upon low-powered, mobile and inexpensive devices - the presence of faulty components and intermittent network connectivity, especially in emergency scenarios, tend to deliver inaccurate/delayed information. The MINES associate team addresses the resulting challenge of enabling interoperability and resilience in large-scale IoT systems through the design and development of a dedicated middleware. More specifically, focusing on emergency situations, the MINES middleware will: (i) enable the dynamic composition of IoT systems from any and all available heterogeneous devices; (ii) support the timely and reliable exchange of critical data within and across IoT in the enabled large-scale and dynamic system over heterogeneous networks. Finally, the team will evaluate the proposed solution in the context of emergency response scenario use cases.

### *9.3.2. Inria Associate Teams Not Involved in an Inria International Lab*

#### *9.3.2.1. HOMENET*

Title: Home network diagnosis and security

International Partner:

Princeton University (United States) - Computer Science Department - Nick Feamster

Start year: 2017

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

Modern households connect a multitude of networked devices (ranging from laptops and smartphones to a number of Internet of Things devices) via a home network. Most home networks, however, do not have a technically skilled network administrator for managing the network, for example to identify faulty equipment or take steps to secure end hosts such as applying security patches. Home networks represent a particularly challenging environment due to the diversity of devices, applications, and services users may connect. The goal of HOMENET is to assist users in diagnosing and securing their home networks. Our approach is based on developing new algorithms and mechanisms that will run on the home router (or in-collaboration with the router). The router connects the home network to the rest of the Internet; it is hence the ideal place to secure home devices and to distinguish problems that happen in the home from those happening elsewhere. We will address a number of research challenges for example in device discovery and fingerprinting, anomaly detection in the Internet of Things, home network diagnosis (including wireless diagnosis). HOMENET will bring together two leading research teams in the network measurement arena with successful prior collaboration. Moreover, Princeton brings an existing home router platform and expertise in security, wireless, and software-defined networks; and MiMove brings an existing Web-based measurement platform, and expertise in traffic-based profiling and anomaly detection.

*9.3.2.2. ACHOR*

Title: Adaptive enactment of service choreographies

International Partner:

Universidade Federal de Goiás (Brazil) - Computer Science Department - Fabio Costa

Start year: 2016

See also: http://www.inf.ufg.br/projects/achor

Service choreographies are distributed compositions of services (e.g., Web services) that coordinate their execution and interactions without centralized control. Due to this decentralized coordination and the ability to compose third-party services, choreographies have shown great potential as an approach to automate the construction of large-scale, on-demand, distributed applications. Technologies to enable this approach are reaching maturity level, such as modeling languages for choreography specification and engines that operate the deployment of services and enactment of choreographies at Future Internet scales. Nevertheless, a number of problems remain open on the way to fully realize the approach, among them: (i) Deployment of multiple choreographies on top of a collection of shared services (considering service sharing as an effective way to increase the utilization of resources); (ii) Dynamic adaptation of functional and non-functional properties due to runtime changes in the environment and user requirements (adapting the set of services and/or the resources used to run the services in order to add/remove/change functions and maintain QoS properties, respectively); and (iii) Seamless and dynamic integration of mobile services (e.g., smartphone apps, sensors and actuators on handhelds and wearables) and cloud- based services (including the need to consider: mobility of both devices and services, resource constraints of mobile devices, temporary disconnection, interoperability between different interaction paradigms (message-passing, event-based, data-sharing) at the middleware layer, and effect of these paradigms on end-to-end QoS). The overall goal of the project is to design an architecture for adaptive middleware to support service choreographies in large-scale scenarios that involve dynamicity and diversity in terms of application requirements, service interaction protocols, and the use of shared local, mobile and cloud resources.

### *9.3.3. Inria International Partners*

*9.3.3.1. Informal International Partners*

- Northeastern University (Prof. David Choffnes): We are working on methods based on active probing to diagnose poor video quality.

## 9.4. International Research Visitors

### *9.4.1. Visits of International Scientists*

Mark Crovella from Boston University is visiting professor at Inria.

### *9.4.2. Visits to International Teams*

*9.4.2.1. Research Stays Abroad*

- Valérie Issarny was visiting scholar at the EECS Department at UC Berkeley till August 2018. She was hosted by CITRIS in the context of which she was carrying out collaborative research in the area of smart cities and acting as scientific coordinator of the Inria@SiliconValley program.
- Renata Teixeira is visiting scholar at the Computer Science department at Stanford University.
- Georgios Bouloukakis is Inria postdoctoral fellow at University of California, Irvine, in the context of the Inria@SiliconValley program.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Selection*

*10.1.1.1. Chair of Conference Program Committees*

- V. Issarny, TPC co-chair of ICSE-SEIS'2018 - Software Engineering in Society Track of the ICSE'18 conference. Gotheborg, SE, May 2018.
- V. Issarny, TPC co-chair of CIC'2018 - 4th IEEE International Conference on Collaboration and Internet Computing. Philadelphia, USA, October 2018
- V. Issarny, TPC co-chair of SCF-ICIOT - 2019 International Conference on Internet of Things Services at the Services Conference Federation. San Diego, USA, June 2019.

*10.1.1.2. Member of the Conference Program Committees*

- V. Christophides, PC member of IEEE Data Engineering Conference 2018.
- V. Issarny, PC member of COOPIS'18, FASE'18 & 20, IC2E'19, ICDCS'18, ICSE'18 & 20, ICSE-SEIS-19, IOTDI'19, Middleware'18, OPODIS'18, WWW'19.
- R. Teixeira, PC member of ACM SIGCOMM 2018.
- R. Teixeira, PC member of ACM HotNets workshop 2018, ACM SIGCOMM Workshop on Self-Driving Networks 2018.
- N. Georgantas is PC member of the following international conferences: SAC'18&'19, AmI'18, SOSE'18&'19, WWW'19.
- N. Georgantas is PC member of the following international workshops: MRT'18, SERENE'18, IoT-ASAP'18&'19, ARM'18, SESOS-ICSE'18.

### *10.1.2. Journal*

*10.1.2.1. Member of the Editorial Boards*

- V. Issarny, Associate editor, ACM Transactions on Internet of Things (TIOT)
- V. Issarny, Associate editor, IEEE Transactions on Services Computing (TSC)
- V. Issarny, Associate editor, IEEE Transactions on Software Engineering (TSE)
- N. Georgantas, Associate editor, International Journal of Ambient Computing and Intelligence (IJACI)
- V. Christophides, Associate editor, MDPI Open Access Journal of Internet of Things (IoT)
- V. Christophides, Associate editor, MDPI Open Access Journal of Future Internet

*10.1.2.2. Reviewer - Reviewing Activities*

- V. Christophides, Reviewer for the ACM Journal Transactions on Internet of Things (TIOT)
- V. Christophides, Reviewer for the Springer Journal Knowledge and Information Systems (KAIS)

### *10.1.3. Invited Talks*

- V. Christophides "IoT Data Analytics", Invited Tutorial at the French Database Community (BDA) Summer School on Scalable Data Analytics Aussois, Modane June 17-22, 2018.
- V. Issarny, "USNB - Enabling Universal Online Social Interactions", Keynote, ECSCW - The 16th European Conference on Computer-Supported Cooperative Work: The International venue on Practice-centred computing and the Design of cooperation technologies. Nancy, FR, June 2018.
- V. Issarny, "When service-oriented and cloud computing meets the IoT: A use case in the context of urban mobile crowd-sensing", Keynote, ESOCC Conference. COMO, IT, September 2018.

- R. Teixeira, “Diagnosis of Internet Quality of Experience in Home Networks”, Keynote, 6th International Conference, NETYS 2018, Essaouira, Morocco, May 9–11, 2018.
- N. Georgantas, "Challenges in Enabling Effective Smart Cities and Communities: Global Perspectives", Plenary panel, IEEE SMARTCOMP 2018, Taormina, Sicily, June 18-20, 2018.

### *10.1.4. Leadership within the Scientific Community*

- V. Christophides, Member of the EDBT Association (since 2014).
- V. Issarny, Council member, ACM Europe (since 2017).

### *10.1.5. Scientific Expertise*

- V. Issarny, Elected member: *Commission d'Evaluation Inria*.
- V. Issarny, Scientific council member: ARCEP (since September 2018).
- V. Issarny, Committee member: *Advanced Data science Alliance Expert Panel of the Canadian Networks of Centres of Excellence* (NCE) (Ottawa, Canada, 08/18); *JPI Urban Europe and NSFC China pilot call on Sustainable and Liveable Cities and Urban Areas* (Paris, 10/18); *FWO Expert panel for PhD fellowships strategic basic research* (Bruxelles, 11/18)
- R. Teixeira, evaluator for the H2020-ICT-2018-2 call, topic ICT-21-2018 “EU-US Collaboration for advanced wireless platforms” organised by the European Commission.
- R. Teixeira, monitor for the H2020 projects CogNet, SUPERFLUIDITY.
- R. Teixeira, Member of the acceptability committee for the competitive selection of young researchers (CR), 2018.
- R. Teixeira, Member of the admission committee for the competitive selection of young researchers (CR), 2018.
- V. Christophides, Member of the appointment committee for faculty position on «Machine Learning with focus on Bioinformatics» University of Crete, December 2017.
- N. Georgantas, monitor for the ANR project INTEROP.

### *10.1.6. Research Administration*

- V. Issarny, Scientific coordinator: IIL Inria@SiliconValley (till 08/18) and IPL CityLab@Inria (till 11/18).
- N. Georgantas, member of the PhD monitoring committee at Inria Paris (till 10/18).

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Master: V. Christophides, "Big Data Processing and Analytics", 25h M2, Ecole CentraleSupelec, France. Created this new master's level class in Spring 2016.

Master : R. Teixeira, “Methodology for research in networking”, 4h CM, M2, Sorbonne University, France

Master : R. Teixeira, “Network Metrology”, 4h CM, M2, Sorbonne University, France

**E-learning**

V. Issarny, MOOC & SPOC *Implementing Successful Processes for Citizen Participation: Supporting Methods and Civic Tech*, 5 weeks, 1,600 registered students in the session of November 2018.

R. Teixeira, MOOC “Internet Measurements: a Hands-on Introduction”, 5 weeks in FUN platform, session May 2018.

### *10.2.2. Supervision*

PhDs in progress:

- Yifan Du (Since October 2017): "In-network collaborative crowd-Xing", Sorbonne University, V. Issarny and F. Sailhan (CNAM)
- Thibaut Ehlinger (Since November 2017): «Causal Modelling of User Quality Experience (QoE) from Network Quality of Service (QoS)» University of Côte d'Azur. advisors: V. Christophides and C. Barakat (Inria)
- Grigoris Piperagkas (Since October 2018): "Leveraging universal social networking and the IoT for urban-scale participatory systems", Sorbone University, V. Issarny and R. Angarita (ISEP).
- Israel Marquez Salinas (Since October 2018): "Optimization of Internet Quality of Experience in Home Networks", S. Ayoubi, R. Teixeira.
- Dimitris Tsolovos (Since March 2017): "A privacy-by-design middleware for urban-scale mobile crowdsensing", UVSQ, N. Anciaux (Inria PETRUS @ Saclay) and V. Issarny.
- Fethi Dilmi (Since October 2017): "End-to-end monitoring and diagnosis of video Quality of Experience", F. Bronzino and R. Teixeira.
- Sarah Wassermann (Since October 2017): "Passive analysis and opitimization of Internet Quality of Experience", R. Teixeira.
- Sara El Aouad (Since May 2013): "Building a personalized summary from movie reviews", V. Christophides, R. Teixeira, C. Diot (Google).
- Patient Ntumba (Since August 2018): "Dynamic adaptation of middleware-layer protocols for emergent mobile systems", Sorbonne University, N. Georgantas.

### *10.2.3. Juries*

V. Christophides: External evaluator of the Ph.D. Thesis of Giorgos Borboudakis entitled "Efficient and Accurate Feature Selection, with Extensions for Multiple Solutions and to Big Data", University of Crete, Greece, November 2018

V. Issarny: President of the PhD defense committee of Saad El Jaouhari on the topic "A secure design of WoT services for smart cities", IMT-Atlantique, Rennes, FR, December 2018

R. Teixeira, external evaluator of the Ph.D dissertation of Danilo Cicalese on "Anycast nowadays" (Telecom ParisTech, 2018).

R. Teixeira, external evaluator of the Ph.D dissertation of Luca Vassio on "Data Analysis and Modelling of Users' Behaviour on the Web" (Politecnico di Torino, 2018).

## 10.3. Popularization

### *10.3.1. Internal or external Inria responsibilities*

- V. Issarny, Appointed member: CODEV, *Métropole du Grand Paris* (Since December 2018).

# 11. Bibliography

## Major publications by the team in recent years

[1] A. BENNACEUR, V. ISSARNY.*Automated Synthesis of Mediators to Support Component Interoperability*, in "IEEE Transactions on Software Engineering", 2015, 22, https://hal.inria.fr/hal-01076176

[2] B. BILLET, V. ISSARNY.*Spinel: An Opportunistic Proxy for Connecting Sensors to the Internet of Things*, in "ACM Transactions on Internet Technology", March 2017, vol. 17, n$^{o}$ 2, p. 1 - 21 [*DOI :* 10.1145/3041025], https://hal.inria.fr/hal-01505879

[3] G. BLAIR, A. BENNACEUR, N. GEORGANTAS, P. GRACE, V. ISSARNY, V. NUNDLOLL, M. PAOLUCCI.*The Role of Ontologies in Emergent Middleware: Supporting Interoperability in Complex Distributed Systems*, in "Big Ideas track of ACM/IFIP/USENIX 12th International Middleware Conference", Lisbon, Portugal, 2011, http://hal.inria.fr/inria-00629059/en

[4] M. CAPORUSCIO, P.-G. RAVERDY, V. ISSARNY.*ubiSOAP: A Service Oriented Middleware for Ubiquitous Networking*, in "IEEE Transactions on Services Computing", 2012, vol. 99 [*DOI :* 10.1109/TSC.2010.60], http://hal.inria.fr/inria-00519577

[5] I. CUNHA, R. TEIXEIRA, D. VEITCH, C. DIOT.*DTrack: A System to Predict and Track Internet Path Changes*, in "IEEE/ACM Transactions on Networking", August 2014, vol. 22, n$^{o}$ 4, p. 1025–1038, https://hal.inria.fr/hal-01097439

[6] O. GOGA, P. LOISEAU, R. SOMMER, R. TEIXEIRA, K. P. GUMMADI.*On the Reliability of Profile Matching Across Large Online Social Networks*, in "KDD'15: ACM SIGDD Conference on Knowledge Discovery and Data Mining", Sydeny, Australia, August 2015 [*DOI :* 10.1145/2783258.2788601], https://hal.inria.fr/hal-01162402

[7] S. HACHEM, A. PATHAK, V. ISSARNY.*Service-Oriented Middleware for Large-Scale Mobile Participatory Sensing*, in "Pervasive and Mobile Computing", 2014, http://hal.inria.fr/hal-00872407

[8] 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 [*DOI :* 10.1007/s13174-011-0021-3], http://hal.inria.fr/inria-00588753/en

[9] K. MIRYLENKA, V. CHRISTOPHIDES, T. PALPANAS, I. PEFKIANAKIS, M. MAY.*Characterizing Home Device Usage From Wireless Traffic Time Series*, in "19th International Conference on Extending Database Technology (EDBT)", Bordeaux, France, March 2016, https://hal.inria.fr/hal-01249778

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[10] N. GEORGANTAS.*Service Oriented Computing in Mobile Environments: Abstractions and Mechanisms for Interoperability and Composition*, Sorbonne Université, February 2018, Habilitation à diriger des recherches, https://hal.inria.fr/tel-01740629

## Articles in International Peer-Reviewed Journal

[11] G. BAJAJ, R. AGARWAL, P. SINGH, N. GEORGANTAS, V. ISSARNY.*4W1H in IoT semantics*, in "IEEE Access", October 2018, vol. 6, p. 65488-65506 [*DOI :* 10.1109/ACCESS.2018.2878100], https://hal.inria.fr/hal-01898506

[12] J. LANZA, L. SANCHEZ, J. R. SANTANA, R. AGARWAL, N. KEFALAKIS, P. GRACE, T. ELSALEH, M. ZHAO, E. TRAGOS, H. NGUYEN, F. CIRILLO, R. STEINKE, J. SOLDATOS.*Experimentation as a Service over Semantically Interoperable Internet of Things Testbeds*, in "IEEE Access", September 2018, vol. 18, n$^{o}$ 1, p. 51607 - 51625 [*DOI :* 10.1109/ACCESS.2018.2867452], https://hal.inria.fr/hal-01876881

[13] L. SANCHEZ, J. LANZA, J. R. SANTANA, R. AGARWAL, P. G. RAVERDY, T. ELSALEH, Y. FATHY, S. JEONG, A. DADOUKIS, T. KORAKIS, S. KERANIDIS, P. O'BRIEN, J. HORGAN, A. SACCHETTI, G. MASTANDREA, A. FRAGKIADAKIS, P. CHARALAMPIDIS, N. SEYDOUX, C. ECREPONT, M. ZHAO.*Federation of Internet of Things Testbeds for the Realization of a Semantically-Enabled Multi-Domain Data Marketplace*, in "Sensors", October 2018, vol. 18, n° 10, 34 [*DOI :* 10.3390/s18103375], https://hal.inria.fr/hal-01889896

[14] I. TSAMARDINOS, G. BORBOUDAKIS, P. KATSOGRIDAKIS, P. PRATIKAKIS, V. CHRISTOPHIDES.*A greedy feature selection algorithm for Big Data of high dimensionality*, in "Machine Learning Journal", August 2018, https://hal.inria.fr/hal-01960888

[15] R. VENTURA, V. MALLET, V. ISSARNY.*Assimilation of mobile phone measurements for noise mapping of a neighborhood*, in "Journal of the Acoustical Society of America", September 2018, vol. 144, n° 3, p. 1279 - 1292 [*DOI :* 10.1121/1.5052173], https://hal.inria.fr/hal-01909933

### Articles in Non Peer-Reviewed Journal

[16] V. MALLET, R. VENTURA, V. ISSARNY, P. G. RAVERDY.*Assimilation d'observations participatives issues de l'application mobile Ambiciti*, in "Acoustique et Techniques : trimestriel d'information des professionnels de l'acoustique", January 2018, https://hal.inria.fr/hal-01909880

### Invited Conferences

[17] V. ISSARNY, G. BOULOUKAKIS, N. GEORGANTAS, F. SAILHAN, G. TEXIER.*When Service-oriented Computing Meets the IoT: A Use Case in the Context of Urban Mobile Crowdsensing*, in "ESOCC 2018 - 7th European Conference on Service-oriented and Cloud Computing", Como, Italy, September 2018, https://hal.inria.fr/hal-01871515

[18] B. LEFÈVRE, V. ISSARNY.*Matching Technological & Societal Innovations: The Social Design of a Mobile Collaborative App for Urban Noise Monitoring*, in "4th IEEE International Conference on Smart Computing", Taormina, Italy, June 2018, https://hal.inria.fr/hal-01801314

### International Conferences with Proceedings

[19] K. E. BENSON, G. BOULOUKAKIS, C. GRANT, V. ISSARNY, S. MEHROTRA, I. MOSCHOLIOS, N. VENKATASUBRAMANIAN.*FireDeX: a Prioritized IoT Data Exchange Middleware for Emergency Response*, in "ACM/IFIP/USENIX Middleware conference", Rennes, France, December 2018, vol. 18 [*DOI :* 10.1145/3274808.3274830], https://hal.inria.fr/hal-01877555

[20] G. BOULOUKAKIS, A. KATTEPUR, N. GEORGANTAS, V. ISSARNY.*Queueing Network Modeling Patterns for Reliable and Unreliable Publish/Subscribe Protocols*, in "MobiQuitous 2018 - 15th EAI International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services", New York, United States, November 2018, https://hal.inria.fr/hal-01893926

[21] G. BOULOUKAKIS, I. MOSCHOLIOS, N. GEORGANTAS, V. ISSARNY.*Simulation-based Queueing Models for Performance Analysis of IoT Applications*, in "11th International Symposium on Communication Systems, Networks, and Digital Signal Processing (CSNDSP)", Budapest, Hungary, July 2018, https://hal.inria.fr/hal-01797930

[22] F. M. COSTA, N. GEORGANTAS, R. GOMES, R. DA ROCHA, G. BOULOUKAKIS.*Cross-layer QoS-Aware Resource Allocation for IoT-Enabled Service Choreographies*, in "The 5th International Workshop on

Middleware and Applications for the Internet of Things (M4IoT)", Rennes, France, December 2018, https://hal.inria.fr/hal-01943541

[23] D. DA HORA, A. S. ASRESE, V. CHRISTOPHIDES, R. TEIXEIRA, D. ROSSI.*Narrowing the gap between QoS metrics and Web QoE using Above-the-fold metrics*, in "PAM 2018 - International Conference on Passive and Active Network Measurement", Berlin, Germany, March 2018, p. 1-13, https://hal.inria.fr/hal-01677260

[24] D. DA HORA, K. VAN DOORSELAER, K. VAN OOST, R. TEIXEIRA.*Predicting the effect of home Wi-Fi quality on QoE*, in "INFOCOM 2018 - IEEE International Conference on Computer Communications", Honolulu, United States, April 2018, p. 1-10, https://hal.inria.fr/hal-01677214

[25] V. EFTHYMIOU, G. PAPADAKIS, K. STEFANIDIS, V. CHRISTOPHIDES.*Simplifying Entity Resolution on Web Data with Schema-agnostic, Non-iterative Matching*, in "ICDE 2018 - 34th IEEE International Conference on Data Engineering", Paris, France, April 2018, p. 1-4, https://hal.inria.fr/hal-01718040

[26] V. EFTHYMIOU, G. PAPADAKIS, K. STEFANIDIS, V. CHRISTOPHIDES.*MinoanER: Schema-Agnostic, Non-Iterative, Massively Parallel Resolution of Web Entities*, in "EDBT/ICDT 2019 Joint Conference", Lisbon, Portugal, March 2019, https://hal.inria.fr/hal-01960933

[27] *Best Paper*
R. GOMES, G. BOULOUKAKIS, F. COSTA, N. GEORGANTAS, R. DA ROCHA.*QoS-Aware Resource Allocation for Mobile IoT Pub/Sub Systems*, in "2018 International Conference on Internet of Things (ICIOT)", Seattle, United States, June 2018, https://hal.inria.fr/hal-01797933.

[28] P. NTUMBA, G. BOULOUKAKIS, N. GEORGANTAS.*Interconnecting and Monitoring Heterogeneous Things in IoT Applications*, in "International Conference on Web Engineering (ICWE)", Caceres, Spain, June 2018, https://hal.inria.fr/hal-01771857

[29] G. TEXIER, V. ISSARNY.*Leveraging the Power of the Crowd and Offloading Urban IoT Networks to Extend their Lifetime*, in "LANMAN 2018 - IEEE International Symposium on Local and Metropolitan Area Networks", Washington, DC, United States, June 2018, p. 1-6, https://hal-imt-atlantique.archives-ouvertes.fr/hal-01813313

### National Conferences with Proceeding

[30] D. TSOLOVOS, N. ANCIAUX, V. ISSARNY.*A Privacy Aware Approach for Participatory Sensing Systems*, in "34ème Conférence sur la Gestion de Données – Principes, Technologies et Applications", Bucharest, Romania, October 2018, https://hal.inria.fr/hal-01947863

### Research Reports

[31] D. DA HORA, K. VAN DOORSELAER, K. VAN OOST, R. TEIXEIRA.*Predicting the effect of home Wi-Fi quality on QoE: Extended Technical Report*, Inria ; Technicolor ; Telecom ParisTech, January 2018, https://hal.inria.fr/hal-01676921

### Other Publications

[32] R. AGARWAL, N. GEORGANTAS, V. ISSARNY.*Impact of Spatial Scale on Mobility Parameters*, June 2018, NetSci 2018 - International Conference on Network Science, Poster, https://hal.inria.fr/hal-01923856

[33] D. DA HORA, D. ROSSI, V. CHRISTOPHIDES, R. TEIXEIRA.*A practical method for measuring Web above-the-fold time*, August 2018, ACM SIGCOMM 2018 Conference, Poster, https://hal.inria.fr/hal-01960984

[34] I. MORANDI.*Service traceroute: Tracing Paths of Application Flows*, UPMC - Paris 6 Sorbonne Universités, September 2018, https://hal.inria.fr/hal-01888618

[35] M.-I. SALINAS.*Home Wi-Fi Impairments*, UPMC (Paris 6) - Sorbonne Université, September 2018, https://hal.inria.fr/hal-01889806

[36] L. SCALZOTTO, K. E. BENSON, G. BOULOUKAKIS, P. BELLAVISTA, V. ISSARNY, S. MEHROTRA, N. VENKATASUBRAMANIAN.*An Implementation Experience with SDN-enabled IoT Data Exchange Middleware*, December 2018, Middleware 2018 - ACM/IFIP/USENIX Middleware conference, Poster, https://hal.inria.fr/hal-01895274

# Project-Team MOKAPLAN

## Advances in Numerical Calculus of Variations

IN COLLABORATION WITH: CEREMADE

IN PARTNERSHIP WITH:
**CNRS**

**Université Paris-Dauphine**

RESEARCH CENTER
**Paris**

THEME
**Numerical schemes and simulations**

# Table of contents

# Project-Team MOKAPLAN

*Creation of the Team: 2013 January 01, updated into Project-Team: 2015 December 01*

**Keywords:**

**Computer Science and Digital Science:**
A5.3. - Image processing and analysis
A5.9. - Signal processing
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.6. - Optimization

**Other Research Topics and Application Domains:**
B1.2. - Neuroscience and cognitive science
B9.5.2. - Mathematics
B9.5.3. - Physics
B9.5.4. - Chemistry
B9.6.3. - Economy, Finance

# 1. Team, Visitors, External Collaborators

**Research Scientists**
Jean-David Benamou [Team leader, Inria, Senior Researcher, HDR]
Yohann de Castro [Univ Paris-Saclay and Ecole Ponts ParisTech, Researcher]
Vincent Duval [Inria, Researcher]
Thomas Gallouët [Inria, Researcher]
Irene Waldspurger [Univ Paris Dauphine, Researcher]

**Faculty Members**
François-Xavier Vialard [Univ Paris Dauphine, Associate Professor]
Guillaume Carlier [Univ Paris Dauphine, Professor, HDR]
Paul Pegon [Univ Paris Dauphine, Assistant Professor, from Oct 2018]

**External Collaborators**
Yann Brenier [Ecole Normale Supérieure Paris, Senior Researcher]
Quentin Mérigot [Univ Paris Saclay, Professor]
Gabriel Peyré [Ecole Normale Supérieure Paris, Senior Researcher, HDR]
Shuangjian Zhang [Ecole Normale Supérieure Paris, Post-Doc, from Jun 2018]

**PhD Students**
Paul Catala [Ecole Normale Supérieure Paris]
Quentin Denoyelle [Univ Paris Dauphine]
Aude Genevay [Univ Paris Dauphine]
Lucas Martinet [Univ Paris Dauphine]
Marco Masoero [Univ Paris Dauphine]
Quentin Ismael Petit [Univ Paris Dauphine, from Sep 2018]
Giorgi Rukhaia [Inria, from May 2018]
Gabriele Todeschi [Univ Paris Dauphine, from Oct 2018]
Miao Yu [Univ Denis Diderot]
Gwendoline de Bie [Ecole Normale Supérieure Paris]

**Post-Doctoral Fellows**

Jean Baptiste Courbot [Univ Paris Sciences et Lettres, until Oct 2018]
Guillaume Garrigos [CNRS, until Sep 2018]
Guillaume Mijoule [Inria, from Jul 2018]
Andrea Natale [Inria]

**Administrative Assistants**

Derya Gok [Inria, from Sep 2018]
Martine Verneuille [Inria, until Sep 2018]

# 2. Overall Objectives

## 2.1. Introduction

The last decade has witnessed a remarkable convergence between several sub-domains of the calculus of variations, namely optimal transport (and its many generalizations), infinite dimensional geometry of diffeomorphisms groups and inverse problems in imaging (in particular sparsity-based regularization). This convergence is due to (i) the mathematical objects manipulated in these problems, namely sparse measures (e.g. coupling in transport, edge location in imaging, displacement fields for diffeomorphisms) and (ii) the use of similar numerical tools from non-smooth optimization and geometric discretization schemes. Optimal Transportation, diffeomorphisms and sparsity-based methods are powerful modeling tools, that impact a rapidly expanding list of scientific applications and call for efficient numerical strategies. Our research program shows the important part played by the team members in the development of these numerical methods and their application to challenging problems.

## 2.2. Static Optimal Transport and Generalizations

### *2.2.1. Optimal Transport, Old and New.*

*Optimal Mass Transportation* is a mathematical research topic which started two centuries ago with Monge's work on the "Théorie des déblais et des remblais" (see [121]). This engineering problem consists in minimizing the transport cost between two given mass densities. In the 40's, Kantorovich [128] introduced a powerful linear relaxation and introduced its dual formulation. The *Monge-Kantorovich* problem became a specialized research topic in optimization and Kantorovich obtained the 1975 Nobel prize in economics for his contributions to resource allocations problems. Since the seminal discoveries of Brenier in the 90's [73], Optimal Transportation has received renewed attention from mathematical analysts and the Fields Medal awarded in 2010 to C. Villani, who gave important contributions to Optimal Transportation and wrote the modern reference monographs [168], [167], arrived at a culminating moment for this theory. Optimal Mass Transportation is today a mature area of mathematical analysis with a constantly growing range of applications. Optimal Transportation has also received a lot of attention from probabilists (see for instance the recent survey [138] for an overview of the Schrödinger problem which is a stochastic variant of the Benamou-Brenier dynamical formulation of optimal transport). The development of numerical methods for Optimal Transportation and Optimal Transportation related problems is a difficult topic and comparatively underdeveloped. This research field has experienced a surge of activity in the last five years, with important contributions of the MOKAPLAN group (see the list of important publications of the team). We describe below a few of recent and less recent Optimal Transportation concepts and methods which are connected to the future activities of MOKAPLAN :

Brenier's theorem [74] characterizes the unique optimal map as the gradient of a convex potential. As such Optimal Transportation may be interpreted as an infinite dimensional optimisation problem under "convexity constraint": i.e. the solution of this infinite dimensional optimisation problem is a convex potential. This connects Optimal Transportation to "convexity constrained" non-linear variational problems such as, for instance, Newton's problem of the body of minimal resistance. The value function of the optimal transport problem is also known to define a distance between source and target densities called the *Wasserstein distance* which plays a key role in many applications such as image processing.

### *2.2.2. Monge-Ampère Methods.*

A formal substitution of the optimal transport map as the gradient of a convex potential in the mass conservation constraint (a Jacobian equation) gives a non-linear Monge-Ampère equation. Caffarelli [83] used this result to extend the regularity theory for the Monge-Ampère equation. In the last ten years, it also motivated new research on numerical solvers for non-linear degenerate Elliptic equations [109] [136] [58] [59] and the references therein. Geometric approaches based on Laguerre diagrams and discrete data [145] have also been developed. Monge-Ampère based Optimal Transportation solvers have recently given the first linear cost computations of Optimal Transportation (smooth) maps.

### *2.2.3. Generalizations of OT.*

In recent years, the classical Optimal Transportation problem has been extended in several directions. First, different ground costs measuring the “physical" displacement have been considered. In particular, well posedness for a large class of convex and concave costs has been established by McCann and Gangbo [120]. Optimal Transportation techniques have been applied for example to a Coulomb ground cost in Quantum chemistry in relation with Density Functional theory [105]. Given the densities of electrons Optimal Transportation models the potential energy and their relative positions. For more than more than 2 electrons (and therefore more than 2 densities) the natural extension of Optimal Transportation is the so called Multi-marginal Optimal Transport (see [149] and the references therein). Another instance of multi-marginal Optimal Transportation arises in the so-called Wasserstein barycenter problem between an arbitrary number of densities [42]. An interesting overview of this emerging new field of optimal transport and its applications can be found in the recent survey of Ghoussoub and Pass [148].

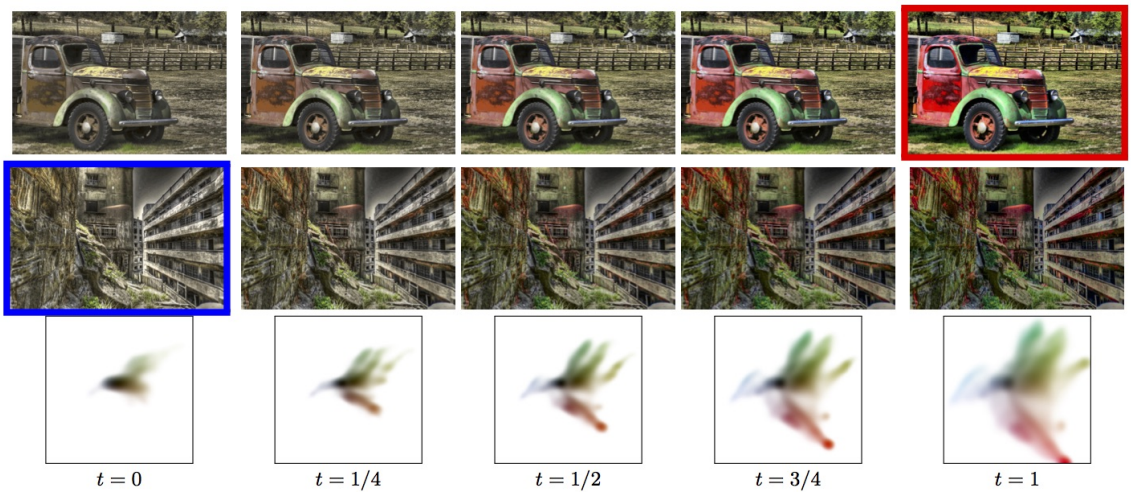


*Figure 1. Example of color transfer between two images, computed using the method developed in [54], see also [161]. The image framed in red and blue are the input images. Top and middle row: adjusted image where the color of the transported histogram has been imposed. Bottom row: geodesic (displacement) interpolation between the histogram of the chrominance of the image.*

### *2.2.4. Numerical Applications of Optimal Transportation.*

Optimal transport has found many applications, starting from its relation with several physical models such as the semi-geostrophic equations in meteorology [125], [107], [106], [51], [135], mesh adaptation [134], the reconstruction of the early mass distribution of the Universe [117], [75] in Astrophysics, and the numerical

optimisation of reflectors following the Optimal Transportation interpretation of Oliker [84] and Wang [169]. Extensions of OT such as multi-marginal transport has potential applications in Density Functional Theory , Generalized solution of Euler equations [72] (DFT) and in statistics and finance [48], [119] .... Recently, there has been a spread of interest in applications of OT methods in imaging sciences [66], statistics [63] and machine learning [108]. This is largely due to the emergence of fast numerical schemes to approximate the transportation distance and its generalizations, see for instance [54]. Figure 1 shows an example of application of OT to color transfer. Figure 9 shows an example of application in computer graphics to interpolate between input shapes.

## 2.3. Diffeomorphisms and Dynamical Transport

### *2.3.1. Dynamical transport.*

While the optimal transport problem, in its original formulation, is a static problem (no time evolution is considered), it makes sense in many applications to rather consider time evolution. This is relevant for instance in applications to fluid dynamics or in medical images to perform registration of organs and model tumor growth.

In this perspective, the optimal transport in Euclidean space corresponds to an evolution where each particule of mass evolves in straight line. This interpretation corresponds to the *Computational Fluid Dynamic* (CFD) formulation proposed by Brenier and Benamou in [50]. These solutions are time curves in the space of densities and geodesics for the Wasserstein distance. The CFD formulation relaxes the non-linear mass conservation constraint into a time dependent continuity equation, the cost function remains convex but is highly non smooth. A remarkable feature of this dynamical formulation is that it can be re-cast as a convex but non smooth optimization problem. This convex dynamical formulation finds many non-trivial extensions and applications, see for instance [52]. The CFD formulation also appears to be a limit case of *Mean Fields games* (MFGs), a large class of economic models introduced by Lasry and Lions [130] leading to a system coupling an Hamilton-Jacobi with a Fokker-Planck equation. In contrast, the Monge case where the ground cost is the euclidan distance leads to a static system of PDEs [68].

### *2.3.2. Gradient Flows for the Wasserstein Distance.*

Another extension is, instead of considering geodesic for transportation metric (i.e. minimizing the Wasserstein distance to a target measure), to make the density evolve in order to minimize some functional. Computing the steepest descent direction with respect to the Wasserstein distance defines a so-called Wasserstein gradient flow, also known as *JKO gradient flows* after its authors [126]. This is a popular tool to study a large class of non-linear diffusion equations. Two interesting examples are the Keller-Segel system for chemotaxis [127], [100] and a model of congested crowd motion proposed by Maury, Santambrogio and Roudneff-Chupin [141]. From the numerical point of view, these schemes are understood to be the natural analogue of implicit scheme for linear parabolic equations. The resolution is however costly as it involves taking the derivative in the Wasserstein sense of the relevant energy, which in turn requires the resolution of a large scale convex but non-smooth minimization.

### *2.3.3. Geodesic on infinite dimensional Riemannian spaces.*

To tackle more complicated warping problems, such as those encountered in medical image analysis, one unfortunately has to drop the convexity of the functional involved in defining the gradient flow. This gradient flow can either be understood as defining a geodesic on the (infinite dimensional) group of diffeomorphisms [47], or on a (infinite dimensional) space of curves or surfaces [170]. The de-facto standard to define, analyze and compute these geodesics is the "Large Deformation Diffeomorphic Metric Mapping" (LDDMM) framework of Trouvé, Younes, Holm and co-authors [47], [124]. While in the CFD formulation of optimal transport, the metric on infinitesimal deformations is just the $L^2$ norm (measure according to the density being transported), in LDDMM, one needs to use a stronger regularizing metric, such as Sobolev-like norms or reproducing kernel Hilbert spaces (RKHS). This enables a control over the smoothness of the deformation which is crucial for many applications. The price to pay is the need to solve a non-convex optimization

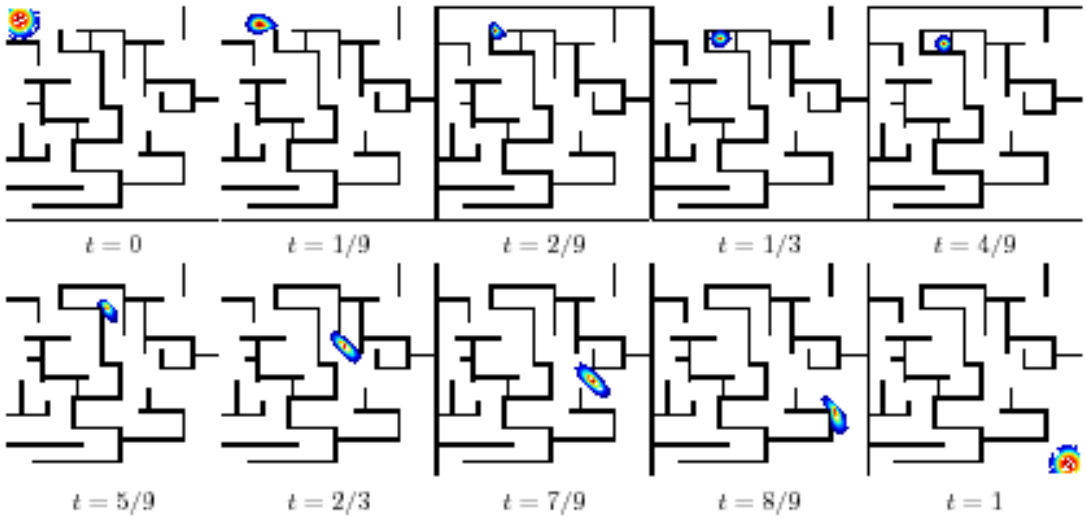


*Figure 2. Examples of displacement interpolation (geodesic for optimal transport) according to a non-Euclidean Riemannian metric (the mass is constrained to move inside a maze) between to input Gaussian distributions. Note that the maze is dynamic: its topology change over time, the mass being "trapped" at time $t = 1/3$.*

problem through geodesic shooting method [142], which requires to integrate backward and forward the geodesic ODE. The resulting strong Riemannian geodesic structure on spaces of diffeomorphisms or shapes is also pivotal to allow us to perform statistical analysis on the tangent space, to define mean shapes and perform dimensionality reduction when analyzing large collection of input shapes (e.g. to study evolution of a diseases in time or the variation across patients) [99].

## 2.4. Sparsity in Imaging

### *2.4.1. Sparse $\ell^1$ regularization.*

Beside image warping and registration in medical image analysis, a key problem in nearly all imaging applications is the reconstruction of high quality data from low resolution observations. This field, commonly referred to as "inverse problems", is very often concerned with the precise location of features such as point sources (modeled as Dirac masses) or sharp contours of objects (modeled as gradients being Dirac masses along curves). The underlying intuition behind these ideas is the so-called sparsity model (either of the data itself, its gradient, or other more complicated representations such as wavelets, curvelets, bandlets [140] and learned representation [171]).

The huge interest in these ideas started mostly from the introduction of convex methods to serve as proxy for these sparse regularizations. The most well known is the $\ell^1$ norm introduced independently in imaging by Donoho and co-workers under the name "Basis Pursuit" [103] and in statistics by Tibshirani [162] under the name "Lasso". A more recent resurgence of this interest dates back to 10 years ago with the introduction of the so-called "compressed sensing" acquisition techniques [88], which make use of randomized forward operators and $\ell^1$-type reconstruction.

### 2.4.2. Regularization over measure spaces.

However, the theoretical analysis of sparse reconstructions involving real-life acquisition operators (such as those found in seismic imaging, neuro-imaging, astro-physical imaging, etc.) is still mostly an open problem. A recent research direction, triggered by a paper of Candès and Fernandez-Granda [87], is to study directly the infinite dimensional problem of reconstruction of sparse measures (i.e. sum of Dirac masses) using the total variation of measures (not to be mistaken for the total variation of 2-D functions). Several works [86], [113], [110] have used this framework to provide theoretical performance guarantees by basically studying how the distance between neighboring spikes impacts noise stability.

Segmentation input output Zooming input output

*Figure 3. Two example of application of the total variation regularization of functions. Left: image segmentation into homogeneous color regions. Right: image zooming (increasing the number of pixels while keeping the edges sharp).*

### 2.4.3. Low complexity regularization and partial smoothness.

In image processing, one of the most popular methods is the total variation regularization [157], [80]. It favors low-complexity images that are piecewise constant, see Figure 3 for some examples on how to solve some image processing problems. Beside applications in image processing, sparsity-related ideas also had a deep impact in statistics [162] and machine learning [44]. As a typical example, for applications to recommendation systems, it makes sense to consider sparsity of the singular values of matrices, which can be relaxed using the so-called nuclear norm (a.k.a. trace norm) [45]. The underlying methodology is to make use of low-complexity regularization models, which turns out to be equivalent to the use of partly-smooth regularization functionals [133], [164] enforcing the solution to belong to a low-dimensional manifold.

## 2.5. Mokaplan unified point of view

The dynamical formulation of optimal transport creates a link between optimal transport and geodesics on diffeomorphisms groups. This formal link has at least two strong implications that MOKAPLAN will elaborate on: (i) the development of novel models that bridge the gap between these two fields ; (ii) the introduction of novel fast numerical solvers based on ideas from both non-smooth optimization techniques and Bregman metrics, as highlighted in Section 3.2.3.

In a similar line of ideas, we believe a unified approach is needed to tackle both sparse regularization in imaging and various generalized OT problems. Both require to solve related non-smooth and large scale optimization problems. Ideas from proximal optimization has proved crucial to address problems in both fields (see for instance [50], [155]). Transportation metrics are also the correct way to compare and regularize variational problems that arise in image processing (see for instance the Radon inversion method proposed in [54]) and machine learning (see [108]). This unity in term of numerical methods is once again at the core of Section 3.2.3.

# 3. Research Program

## 3.1. Modeling and Analysis

The first layer of methodological tools developed by our team is a set of theoretical continuous models that aim at formalizing the problems studied in the applications. These theoretical findings will also pave the way to efficient numerical solvers that are detailed in Section 3.2.

### *3.1.1. Static Optimal Transport and Generalizations*

*3.1.1.1. Convexity constraint and Principal Agent problem in Economics.*

(*Participants:* G. Carlier, J-D. Benamou, V. Duval, Xavier Dupuis (LUISS Guido Carli University, Roma)) The principal agent problem plays a distinguished role in the literature on asymmetric information and contract theory (with important contributions from several Nobel prizes such as Mirrlees, Myerson or Spence) and it has many important applications in optimal taxation, insurance, nonlinear pricing. The typical problem consists in finding a cost minimizing strategy for a monopolist facing a population of agents who have an unobservable characteristic, the principal therefore has to take into account the so-called incentive compatibilty constraint which is very similar to the cyclical monotonicity condition which characterizes optimal transport plans. In a special case, Rochet and Choné [156] reformulated the problem as a variational problem subject to a convexity constraint. For more general models, and using ideas from Optimal Transportation, Carlier [90] considered the more general $c$-convexity constraint and proved a general existence result. Using the formulation of [90] McCann, Figalli and Kim [114] gave conditions under which the principal agent problem can be written as an infinite dimensional convex variational problem. The important results of [114] are intimately connected to the regularity theory for optimal transport and showed that there is some hope to numerically solve the principal-agent problem for general utility functions.
*Our expertise:* We have already contributed to the numerical resolution of the Principal Agent problem in the case of the convexity constraint, see [95], [146], [143].
*Goals:* So far, the mathematical PA model can be numerically solved for simple utility functions. A Bregman approach inspired by [54] is currently being developed [93] for more general functions. It would be extremely useful as a complement to the theoretical analysis. A new semi-Discrete Geometric approach is also investigated where the method reduces to non-convex polynomial optimization.

*3.1.1.2. Optimal transport and conditional constraints in statistics and finance.*

(*Participants:* G. Carlier, J-D. Benamou, G. Peyré) A challenging branch of emerging generalizations of Optimal Transportation arising in *economics, statistics and finance* concerns Optimal Transportation with *conditional* constraints. The *martingale optimal transport* [48], [119] which appears naturally in mathematical finance aims at computing robust bounds on option prices as the value of an optimal transport problem where not only the marginals are fixed but the coupling should be the law of a martingale, since it represents the prices of the underlying asset under the risk-neutral probability at the different dates. Note that as soon as more than two dates are involved, we are facing a multimarginal problem.
*Our expertise:* Our team has a deep expertise on the topic of OT and its generalization, including many already existing collaboration between its members, see for instance [54], [60], [52] for some representative recent collaborative publications.
*Goals:* This is a non trivial extension of Optimal Transportation theory and MOKAPLAN will develop numerical methods (in the spirit of entropic regularization) to address it. A popular problem in statistics is the so-called quantile regression problem, recently Carlier, Chernozhukov and Galichon [91] used an Optimal Transportation approach to extend quantile regression to several dimensions. In this approach again, not only fixed marginals constraints are present but also constraints on conditional means. As in the martingale Optimal Transportation problem, one has to deal with an extra conditional constraint. The duality approach usually breaks down under such constraints and characterization of optimal couplings is a challenging task both from a theoretical and numerical viewpoint.

*3.1.1.3. JKO gradient flows.*

(*Participants:* G. Carlier, J-D. Benamou, M. Laborde, Q. Mérigot, V. Duval) The connection between the static and dynamic transportation problems (see Section 2.3) opens the door to many extensions, most notably by leveraging the use of gradient flows in metric spaces. The flow with respect to the transportation distance has been introduced by Jordan-Kindelherer-Otto (JKO) [126] and provides a variational formulation of many linear and non-linear diffusion equations. The prototypical example is the Fokker Planck equation. We will explore this formalism to study new variational problems over probability spaces, and also to derive innovative numerical solvers. The JKO scheme has been very successfully used to study evolution equations that have the structure of a gradient flow in the Wasserstein space. Indeed many important PDEs have this structure: the Fokker-Planck equation (as was first considered by [126]), the porous medium equations, the granular media equation, just to give a few examples. It also finds application in image processing [79]. Figure 4 shows examples of gradient flows.
*Our expertise:* There is an ongoing collaboration between the team members on the theoretical and numerical analysis of gradient flows.
*Goals:* We apply and extend our research on JKO numerical methods to treat various extensions:

- Wasserstein gradient flows with a non displacement convex energy (as in the parabolic-elliptic Keller-Segel chemotaxis model [97])
- systems of evolution equations which can be written as gradient flows of some energy on a product space (possibly mixing the Wasserstein and $L^2$ structures) : multi-species models or the parabolic-parabolic Keller-Segel model [64]
- perturbation of gradient flows: multi-species or kinetic models are not gradient flows, but may be viewed as a perturbation of Wasserstein gradient flows, we shall therefore investigate convergence of splitting methods for such equations or systems.

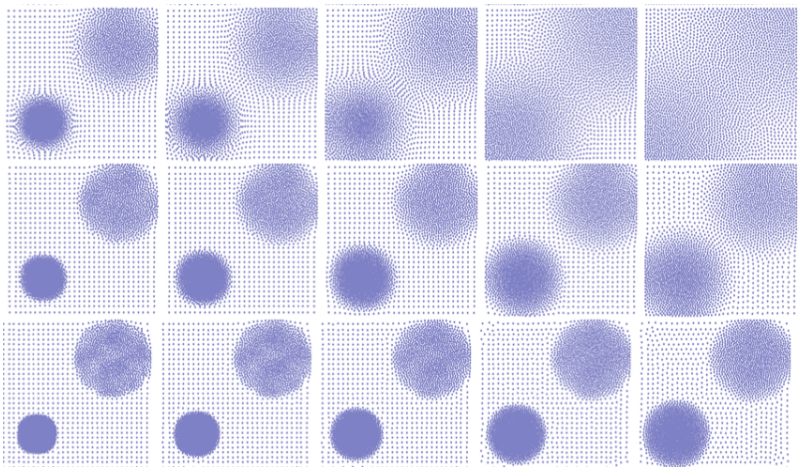

*Figure 4. Example of non-linear diffusion equations solved with a JKO flow [56]. The horizontal axis shows the time evolution minimizing the functional $\int \frac{\rho^\alpha}{\alpha-1}$ on the density $\rho$ (discretized here using point clouds, i.e. sum of Diracs' with equal mass). Each row shows a different value of $\alpha = (0.6, 2, 3)$*

*3.1.1.4. From networks to continuum congestion models.*

(*Participants:* G. Carlier, J-D. Benamou, G. Peyré) Congested transport theory in the discrete framework of networks has received a lot of attention since the 50's starting with the seminal work of Wardrop. A few years later, Beckmann proved that equilibria are characterized as solution of a convex minimization problem. However, this minimization problem involves one flow variable per path on the network, its dimension thus quickly becomes too large in practice. An alternative, is to consider continuous in space models of congested optimal transport as was done in [94] which leads to very degenerate PDEs [70].
*Our expertise:* MOKAPLAN members have contributed a lot to the analysis of congested transport problems and to optimization problems with respect to a metric which can be attacked numerically by fast marching methods [60].
*Goals:* The case of general networks/anisotropies is still not well understood, general $\Gamma$-convergence results will be investigated as well as a detailed analysis of the corresponding PDEs and numerical methods to solve them. Benamou and Carlier already studied numerically some of these PDEs by an augmented Lagrangian method see figure 5. Note that these class of problems share important similarities with metric learning problem in machine learning, detailed below.

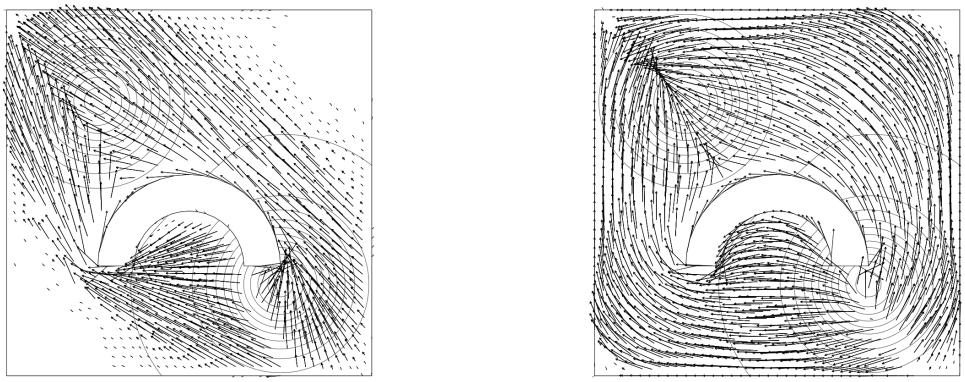

*Figure 5. Monge and Wardrop flows of mass around an obstacle [52]. the source/target mass is represented by the level curves. Left : no congestion, Right : congestion.*

### *3.1.2. Diffeomorphisms and Dynamical Transport*

*3.1.2.1. Growth Models for Dynamical Optimal Transport.*

(*Participants:* F-X. Vialard, J-D. Benamou, G. Peyré, L. Chizat) A major issue with the standard dynamical formulation of OT is that it does not allow for variation of mass during the evolution, which is required when tackling medical imaging applications such as tumor growth modeling [82] or tracking elastic organ movements [160]. Previous attempts [137], [153] to introduce a source term in the evolution typically lead to mass teleportation (propagation of mass with infinite speed), which is not always satisfactory.
*Our expertise:* Our team has already established key contributions both to connect OT to fluid dynamics [50] and to define geodesic metrics on the space of shapes and diffeomorphisms [102].
*Goals:* Lenaic Chizat's PhD thesis aims at bridging the gap between dynamical OT formulation, and LDDDM diffeomorphisms models (see Section 2.3). This will lead to biologically-plausible evolution models that are both more tractable numerically than LDDM competitors, and benefit from strong theoretical guarantees associated to properties of OT.

*3.1.2.2. Mean-field games.*

(*Participants:* G. Carlier, J-D. Benamou) The Optimal Transportation Computational Fluid Dynamics (CFD) formulation is a limit case of variational Mean-Field Games (MFGs), a new branch of game theory recently developed by J-M. Lasry and P-L. Lions [130] with an extremely wide range of potential applications [122]. Non-smooth proximal optimization methods used successfully for the Optimal Transportation can be used in the case of deterministic MFGs with singular data and/or potentials [53]. They provide a robust treatment of the positivity constraint on the density of players.
*Our expertise:* J.-D. Benamou has pioneered with Brenier the CFD approach to Optimal Transportation. Regarding MFGs, on the numerical side, our team has already worked on the use of augmented Lagrangian methods in MFGs [52] and on the analytical side [89] has explored rigorously the optimality system for a singular CFD problem similar to the MFG system.
*Goals:* We will work on the extension to stochastic MFGs. It leads to non-trivial numerical difficulties already pointed out in [41].

*3.1.2.3. Macroscopic Crowd motion, congestion and equilibria.*

(*Participants:* G. Carlier, J-D. Benamou, Q. Mérigot, F. Santambrogio (U. Paris-Sud), Y. Achdou (Univ. Paris 7), R. Andreev (Univ. Paris 7)) Many models from PDEs and fluid mechanics have been used to give a description of *people or vehicles moving in a congested environment*. These models have to be classified according to the dimension (1D model are mostly used for cars on traffic networks, while 2-D models are most suitable for pedestrians), to the congestion effects ("soft" congestion standing for the phenomenon where high densities slow down the movement, "hard" congestion for the sudden effects when contacts occur, or a certain threshold is attained), and to the possible rationality of the agents Maury et al [141] recently developed a theory for 2D hard congestion models without rationality, first in a discrete and then in a continuous framework. This model produces a PDE that is difficult to attack with usual PDE methods, but has been successfully studied via Optimal Transportation techniques again related to the JKO gradient flow paradigm. Another possibility to model crowd motion is to use the mean field game approach of Lions and Lasry which limits of Nash equilibria when the number of players is large. This also gives macroscopic models where congestion may appear but this time a global equilibrium strategy is modelled rather than local optimisation by players like in the JKO approach. Numerical methods are starting to be available, see for instance [41], [78].
*Our expertise:* We have developed numerical methods to tackle both the JKO approach and the MFG approach. The Augmented Lagrangian (proximal) numerical method can actually be applied to both models [52], JKO and deterministic MFGs.
*Goals:* We want to extend our numerical approach to more realistic congestion model where the speed of agents depends on the density, see Figure 6 for preliminary results. Comparison with different numerical approaches will also be performed inside the ANR ISOTACE. Extension of the Augmented Lagrangian approach to Stochastic MFG will be studied.

*3.1.2.4. Diffeomorphic image matching.*

(*Participants:* F-X. Vialard, G. Peyré, B. Schmitzer, L. Chizat) Diffeomorphic image registration is widely used in medical image analysis. This class of problems can be seen as the computation of a generalized optimal transport, where the optimal path is a geodesic on a group of diffeomorphisms. The major difference between the two approaches being that optimal transport leads to non smooth optimal maps in general, which is however compulsory in diffeomorphic image matching. In contrast, optimal transport enjoys a convex variational formulation whereas in LDDMM the minimization problem is non convex.
*Our expertise:* F-X. Vialard is an expert of diffeomorphic image matching (LDDMM) [165], [76], [163]. Our team has already studied flows and geodesics over non-Riemannian shape spaces, which allows for piecewise smooth deformations [102].

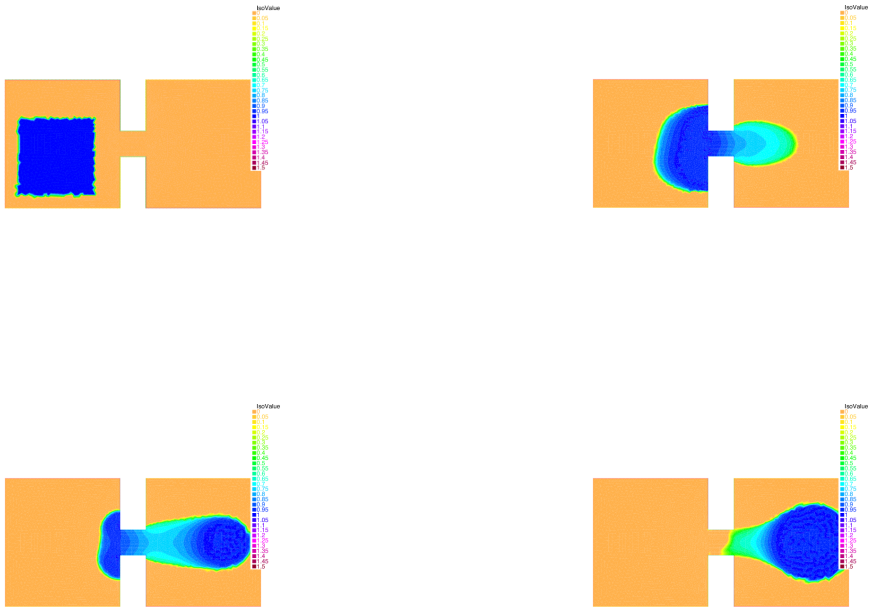

*Figure 6. Example of crowd congestion with density dependent speed. The macroscopic density, at 4 different times, of people forced to exit from one room towards a meeting point in a second room.*

*Goals:* Our aim consists in bridging the gap between standard optimal transport and diffeomorphic methods by building new diffeomorphic matching variational formulations that are convex (geometric obstructions might however appear). A related perspective is the development of new registration/transport models in a Lagrangian framework, in the spirit of [159], [160] to obtain more meaningful statistics on longitudinal studies.

Diffeomorphic matching consists in the minimization of a functional that is a sum of a deformation cost and a similarity measure. The choice of the similarity measure is as important as the deformation cost. It is often chosen as a norm on a Hilbert space such as functions, currents or varifolds. From a Bayesian perspective, these similarity measures are related to the noise model on the observed data which is of geometric nature and it is not taken into account when using Hilbert norms. Optimal transport fidelity have been used in the context of signal and image denoising [132], and it is an important question to extends these approach to registration problems. Therefore, we propose to develop similarity measures that are geometric and computationally very efficient using entropic regularization of optimal transport.

Our approach is to use a regularized optimal transport to design new similarity measures on all of those Hilbert spaces. Understanding the precise connections between the evolution of shapes and probability distributions will be investigated to cross-fertilize both fields by developing novel transportation metrics and diffeomorphic shape flows.

The corresponding numerical schemes are however computationally very costly. Leveraging our understanding of the dynamic optimal transport problem and its numerical resolution, we propose to develop new algorithms. These algorithms will use the smoothness of the Riemannian metric to improve both accuracy and speed, using for instance higher order minimization algorithm on (infinite dimensional) manifolds.

*3.1.2.5. Metric learning and parallel transport for statistical applications.*

(*Participants:* F-X. Vialard, G. Peyré, B. Schmitzer, L. Chizat) The LDDMM framework has been advocated to enable statistics on the space of shapes or images that benefit from the estimation of the deformation. The statistical results of it strongly depend on the choice of the Riemannian metric. A possible direction consists in learning the right invariant Riemannian metric as done in [166] where a correlation matrix (Figure 7) is learnt which represents the covariance matrix of the deformation fields for a given population of shapes. In the same direction, a question of emerging interest in medical imaging is the analysis of time sequence of shapes (called longitudinal analysis) for early diagnosis of disease, for instance [115]. A key question is the inter subject comparison of the organ evolution which is usually done by transport of the time evolution in a common coordinate system via parallel transport or other more basic methods. Once again, the statistical results (Figure 8) strongly depend on the choice of the metric or more generally on the connection that defines parallel transport.
*Our expertise:* Our team has already studied statistics on longitudinal evolutions in [115], [116].
*Goals:* Developing higher order numerical schemes for parallel transport (only low order schemes are available at the moment) and developing variational models to learn the metric or the connections for improving statistical results.

### *3.1.3. Sparsity in Imaging*

*3.1.3.1. Inverse problems over measures spaces.*

(*Participants:* G. Peyré, V. Duval, C. Poon, Q. Denoyelle) As detailed in Section 2.4, popular methods for regularizing inverse problems in imaging make use of variational analysis over infinite-dimensional (typically non-reflexive) Banach spaces, such as Radon measures or bounded variation functions.
*Our expertise:* We have recently shown in [164] how – in the finite dimensional case – the non-smoothness of the functionals at stake is crucial to enforce the emergence of geometrical structures (edges in images or fractures in physical materials [65]) for discrete (finite dimensional) problems. We extended this result in a simple infinite dimensional setting, namely sparse regularization of Radon measures for deconvolution [110]. A deep understanding of those continuous inverse problems is crucial to analyze the behavior of their discrete counterparts, and in [111] we have taken advantage of this understanding to develop a fine analysis of the artifacts induced by discrete (*i.e.* which involve grids) deconvolution models. These works are also closely

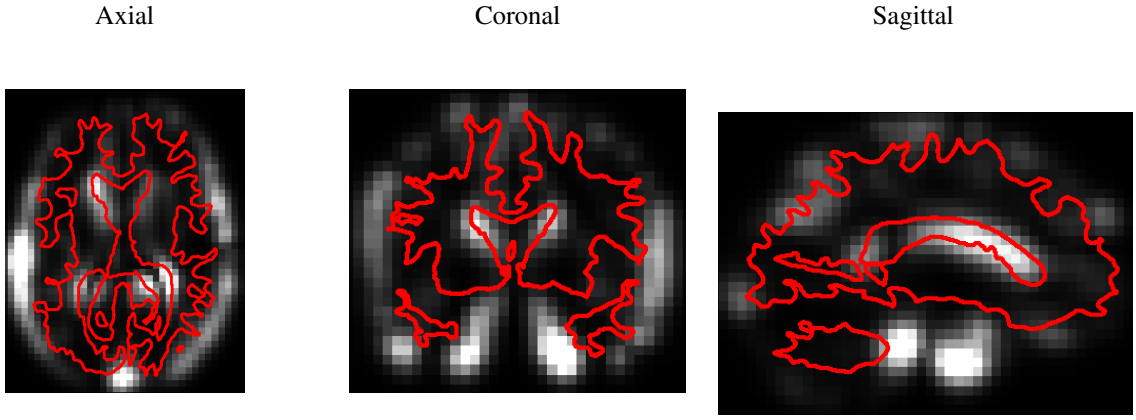


*Figure 7. Learning Riemannian metrics in diffeomorphic image matching to capture the brain variability: a diagonal operator that encodes the Riemannian metric is learnt on a template brain out of a collection of brain images. The values of the diagonal operator are shown in greyscale. The red curves represent the boundary between white and grey matter. For more details, we refer the reader to [166], which was a first step towards designing effective and robust metric learning algorithms.*

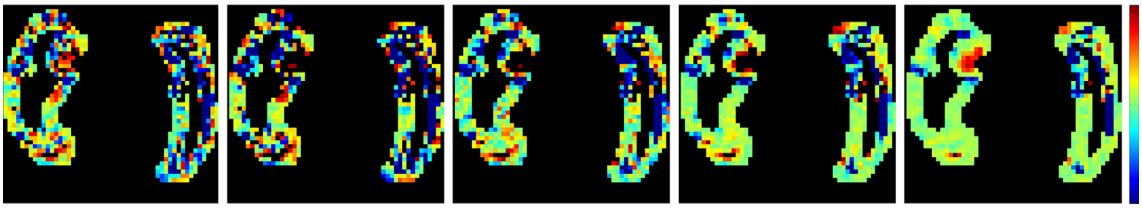

*Figure 8. Statistics on initial momenta: In [115], we compared several intersubject transport methodologies to perform statistics on longitudinal evolutions. These longitudinal evolutions are represented by an initial velocity field on the shapes boundaries and these velocity fields are then compared using logistic regression methods that are regularized. The four pictures represent different regularization methods such as $L^2$, $H^1$ and regularization including a sparsity prior such as Lasso, Fused Lasso and $TV$.*

related to the problem of limit analysis and yield design in mechanical plasticity, see [92], [65] for an existing collaboration between MOKAPLAN's team members.
*Goals:* A current major front of research in the mathematical analysis of inverse problems is to extend these results for more complicated infinite dimensional signal and image models, such as for instance the set of piecewise regular functions. The key bottleneck is that, contrary to sparse measures (which are finite sums of Dirac masses), here the objects to recover (smooth edge curves) are not parameterized by a finite number of degrees of freedom. The relevant previous work in this direction are the fundamental results of Chambolle, Caselles and co-workers [49], [43], [98]. They however only deal with the specific case where there is no degradation operator and no noise in the observations. We believe that adapting these approaches using our construction of vanishing derivative pre-certificate [110] could lead to a solution to these theoretical questions.

*3.1.3.2. Sub-Riemannian diffusions.*

(*Participants:* G. Peyré, J-M. Mirebeau, D. Prandi) Modeling and processing natural images require to take into account their geometry through anisotropic diffusion operators, in order to denoise and enhance directional features such as edges and textures [152], [112]. This requirement is also at the heart of recently proposed models of cortical processing [151]. A mathematical model for these processing is diffusion on sub-Riemanian manifold. These methods assume a fixed, usually linear, mapping from the 2-D image to a lifted function defined on the product of space and orientation (which in turn is equipped with a sub-Riemannian manifold structure).
*Our expertise:* J-M. Mirebeau is an expert in the discretization of highly anisotropic diffusions through the use of locally adaptive computational stencils [144], [112]. G. Peyré has done several contributions on the definition of geometric wavelets transform and directional texture models, see for instance [152]. Dario Prandi has recently applied methods from sub-Riemannian geometry to image restoration [67].
*Goals:* A first aspect of this work is to study non-linear, data-adaptive, lifting from the image to the space/orientation domain. This mapping will be implicitly defined as the solution of a convex variational problem. This will open both theoretical questions (existence of a solution and its geometrical properties, when the image to recover is piecewise regular) and numerical ones (how to provide a faithful discretization and fast second order Newton-like solvers). A second aspect of this task is to study the implication of these models for biological vision, in a collaboration with the UNIC Laboratory (directed by Yves Fregnac), located in Gif-sur-Yvette. In particular, the study of the geometry of singular vectors (or "ground states" using the terminology of [61]) of the non-linear sub-Riemannian diffusion operators is highly relevant from a biological modeling point of view.

*3.1.3.3. Sparse reconstruction from scanner data.*

(*Participants:* G. Peyré, V. Duval, C. Poon) Scanner data acquisition is mathematically modeled as a (sub-sampled) Radon transform [123]. It is a difficult inverse problem because the Radon transform is ill-posed and the set of observations is often aggressively sub-sampled and noisy [158]. Typical approaches [129] try to recover piecewise smooth solutions in order to recover precisely the position of the organ being imaged. There is however a very poor understanding of the actual performance of these methods, and little is known on how to enhance the recovery.
*Our expertise:* We have obtained a good understanding of the performance of inverse problem regularization on *compact* domains for pointwise sources localization [110].
*Goals:* We aim at extending the theoretical performance analysis obtained for sparse measures [110] to the set of piecewise regular 2-D and 3-D functions. Some interesting previous work of C. Poon et al [154] (C. Poon is currently a postdoc in MOKAPLAN) have tackled related questions in the field of variable Fourier sampling for compressed sensing application (which is a toy model for fMRI imaging). These approaches are however not directly applicable to Radon sampling, and require some non-trivial adaptations. We also aim at better exploring the connection of these methods with optimal-transport based fidelity terms such as those introduced in [40].

*3.1.3.4. Tumor growth modeling in medical image analysis.*

(*Participants:* G. Peyré, F-X. Vialard, J-D. Benamou, L. Chizat) Some applications in medical image analysis require to track shapes whose evolution is governed by a growth process. A typical example is tumor growth, where the evolution depends on some typically unknown but meaningful parameters that need to be estimated. There exist well-established mathematical models [82], [150] of non-linear diffusions that take into account recently biologically observed property of tumors. Some related optimal transport models with mass variations have also recently been proposed [139], which are connected to so-called metamorphoses models in the LDDMM framework [62].
*Our expertise:* Our team has a strong experience on both dynamical optimal transport models and diffeomorphic matching methods (see Section 3.1.2).
*Goals:* The close connection between tumor growth models [82], [150] and gradient flows for (possibly non-Euclidean) Wasserstein metrics (see Section 3.1.2) makes the application of the numerical methods we develop particularly appealing to tackle large scale forward tumor evolution simulation. A significant departure from the classical OT-based convex models is however required. The final problem we wish to solve is the backward (inverse) problem of estimating tumor parameters from noisy and partial observations. This also requires to set-up a meaningful and robust data fidelity term, which can be for instance a generalized optimal transport metric.

## 3.2. Numerical Tools

The above continuous models require a careful discretization, so that the fundamental properties of the models are transferred to the discrete setting. Our team aims at developing innovative discretization schemes as well as associated fast numerical solvers, that can deal with the geometric complexity of the variational problems studied in the applications. This will ensure that the discrete solution is correct and converges to the solution of the continuous model within a guaranteed precision. We give below examples for which a careful mathematical analysis of the continuous to discrete model is essential, and where dedicated non-smooth optimization solvers are required.

### *3.2.1. Geometric Discretization Schemes*

*3.2.1.1. Discretizing the cone of convex constraints.*

(*Participants:* J-D. Benamou, G. Carlier, J-M. Mirebeau, Q. Mérigot) Optimal transportation models as well as continuous models in economics can be formulated as infinite dimensional convex variational problems with the constraint that the solution belongs to the cone of convex functions. Discretizing this constraint is however a tricky problem, and usual finite element discretizations fail to converge.
*Our expertise:* Our team is currently investigating new discretizations, see in particular the recent proposal [59] for the Monge-Ampère equation and [143] for general non-linear variational problems. Both offer convergence guarantees and are amenable to fast numerical resolution techniques such as Newton solvers. Since [59] explaining how to treat efficiently and in full generality Transport Boundary Conditions for Monge-Ampère, this is a promising fast and new approach to compute Optimal Transportation viscosity solutions. A monotone scheme is needed. One is based on Froese Oberman work [118], a new different and more accurate approach has been proposed by Mirebeau, Benamou and Collino [57]. As shown in [104], discretizing the constraint for a continuous function to be convex is not trivial. Our group has largely contributed to solve this problem with G. Carlier [95], Quentin Mérigot [146] and J-M. Mirebeau [143]. This problem is connected to the construction of monotone schemes for the Monge-Ampère equation.
*Goals:* The current available methods are 2-D. They need to be optimized and parallelized. A non-trivial extension to 3-D is necessary for many applications. The notion of $c$-convexity appears in optimal transport for generalized displacement costs. How to construct an adapted discretization with "good" numerical properties is however an open problem.

*3.2.1.2. Numerical JKO gradient flows.*

(*Participants:* J-D. Benamou, G. Carlier, J-M. Mirebeau, G. Peyré, Q. Mérigot) As detailed in Section 2.3, gradient Flows for the Wasserstein metric (aka JKO gradient flows [126]) provides a variational formulation of many non-linear diffusion equations. They also open the way to novel discretization schemes. From a computational point, although the JKO scheme is constructive (it is based on the implicit Euler scheme), it has not been very much used in practice numerically because the Wasserstein term is difficult to handle (except in dimension one).
*Our expertise:*

Solving one step of a JKO gradient flow is similar to solving an Optimal transport problem. A geometrical a discretization of the Monge-Ampère operator approach has been proposed by Mérigot, Carlier, Oudet and Benamou in [56] see Figure 4. The Gamma convergence of the discretisation (in space) has been proved.
*Goals:* We are also investigating the application of other numerical approaches to Optimal Transport to JKO gradient flows either based on the CFD formulation or on the entropic regularization of the Monge-Kantorovich problem (see section 3.2.3). An in-depth study and comparison of all these methods will be necessary.

### 3.2.2. Sparse Discretization and Optimization

*3.2.2.1. From discrete to continuous sparse regularization and transport.*

(*Participants:* V. Duval, G. Peyré, G. Carlier, Jalal Fadili (ENSICaen), Jérôme Malick (CNRS, Univ. Grenoble)) While pervasive in the numerical analysis community, the problem of discretization and $\Gamma$-convergence from discrete to continuous is surprisingly over-looked in imaging sciences. To the best of our knowledge, our recent work [110], [111] is the first to give a rigorous answer to the transition from discrete to continuous in the case of the spike deconvolution problem. Similar problems of $\Gamma$-convergence are progressively being investigated in the optimal transport community, see in particular [96].
*Our expertise:* We have provided the first results on the discrete-to-continous convergence in both sparse regularization variational problems [110], [111] and the static formulation of OT and Wasserstein barycenters [96]
*Goals:* In a collaboration with Jérôme Malick (Inria Grenoble), our first goal is to generalize the result of [110] to generic partly-smooth convex regularizers routinely used in imaging science and machine learning, a prototypal example being the nuclear norm (see [164] for a review of this class of functionals). Our second goal is to extend the results of [96] to the novel class of entropic discretization schemes we have proposed [54], to lay out the theoretical foundation of these ground-breaking numerical schemes.

*3.2.2.2. Polynomial optimization for grid-free regularization.*

(*Participants:* G. Peyré, V. Duval, I. Waldspurger) There has been a recent spark of attention of the imaging community on so-called "grid free" methods, where one tries to directly tackle the infinite dimensional recovery problem over the space of measures, see for instance [87], [110]. The general idea is that if the range of the imaging operator is finite dimensional, the associated dual optimization problem is also finite dimensional (for deconvolution, it corresponds to optimization over the set of trigonometric polynomials).
*Our expertise:* We have provided in [110] a sharp analysis of the support recovery property of this class of methods for the case of sparse spikes deconvolution.
*Goals:* A key bottleneck of these approaches is that, while being finite dimensional, the dual problem necessitates to handle a constraint of polynomial positivity, which is notoriously difficult to manipulate (except in the very particular case of 1-D problems, which is the one exposed in [87]). A possible, but very costly, methodology is to ressort to Lasserre's SDP representation hierarchy [131]. We will make use of these approaches and study how restricting the level of the hierarchy (to obtain fast algorithms) impacts the recovery performances (since this corresponds to only computing approximate solutions). We will pay a particular attention to the recovery of 2-D piecewise constant functions (the so-called total variation of functions regularization [157]), see Figure 3 for some illustrative applications of this method.

### *3.2.3. First Order Proximal Schemes*

*3.2.3.1. $L^2$ proximal methods.*

(*Participants:* G. Peyré, J-D. Benamou, G. Carlier, Jalal Fadili (ENSICaen)) Both sparse regularization problems in imaging (see Section 2.4) and dynamical optimal transport (see Section 2.3) are instances of large scale, highly structured, non-smooth convex optimization problems. First order proximal splitting optimization algorithms have recently gained lots of interest for these applications because they are the only ones capable of scaling to giga-pixel discretizations of images and volumes and at the same time handling non-smooth objective functions. They have been successfully applied to optimal transport [50], [147], congested optimal transport [81] and to sparse regularizations (see for instance [155] and the references therein).
*Our expertise:* The pioneering work of our team has shown how these proximal solvers can be used to tackle the dynamical optimal transport problem [50], see also [147]. We have also recently developed new proximal schemes that can cope with non-smooth composite objectives functions [155].
*Goals:* We aim at extending these solvers to a wider class of variational problems, most notably optimization under divergence constraints [52]. Another subject we are investigating is the extension of these solvers to both non-smooth and non-convex objective functionals, which are mandatory to handle more general transportation problems and novel imaging regularization penalties.

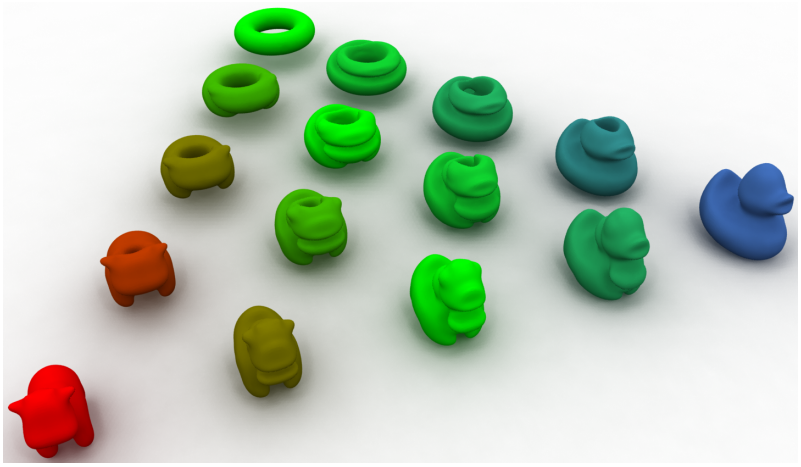

*Figure 9. Example of barycenter between shapes computed using optimal transport barycenters of the uniform densities inside the 3 extremal shapes, computed as detailed in [161]. Note that the barycenters are not in general uniform distributions, and we display them as the surface defined by a suitable level-set of the density.*

*3.2.3.2. Bregman proximal methods.*

(*Participants:* G. Peyré G. Carlier, L. Nenna, J-D. Benamou, L. Nenna, Marco Cuturi (Kyoto Univ.)) The entropic regularization of the Kantorovich linear program for OT has been shown to be surprisingly simple and efficient, in particular for applications in machine learning [108]. As shown in [54], this is a special instance of the general method of Bregman iterations, which is also a particular instance of first order proximal schemes according to the Kullback-Leibler divergence.

*Our expertise:* We have recently [54] shown how Bregman projections [71] and Dykstra algorithm [46] offer a generic optimization framework to solve a variety of generalized OT problems. Carlier and Dupuis [93] have designed a new method based on alternate Dykstra projections and applied it to the *principal-agent problem* in microeconomics. We have applied this method in computer graphics in a paper accepted in SIGGRAPH 2015 [161]. Figure 9 shows the potential of our approach to handle giga-voxel datasets: the input volumetric densities are discretized on a $100^3$ computational grid.
*Goals:* Following some recent works (see in particular [101]) we first aim at studying primal-dual optimization schemes according to Bregman divergences (that would go much beyond gradient descent and iterative projections), in order to offer a versatile and very effective framework to solve variational problems involving OT terms. We then also aim at extending the scope of usage of this method to applications in quantum mechanics (Density Functional Theory, see [105]) and fluid dynamics (Brenier's weak solutions of the incompressible Euler equation, see [72]). The computational challenge is that realistic physical examples are of a huge size not only because of the space discretization of one marginal but also because of the large number of marginals involved (for incompressible Euler the number of marginals equals the number of time steps).

# 4. Highlights of the Year

## 4.1. Highlights of the Year

### *4.1.1. Awards*

G. Carlier was a John von Neumann invited Professor at TUM (Munich) in 2018.

# 5. New Software and Platforms

## 5.1. ALG2

FUNCTIONAL DESCRIPTION: ALG2 for Monge Mean-Field Games, Monge problem and Variational problems under divergence constraint. A generalisation of the ALG2 algorithm has been implemented in FreeFem++.

- Contact: Jean-David Benamou
- URL: https://team.inria.fr/mokaplan/augmented-lagrangian-simulations/

## 5.2. Mokabajour

FUNCTIONAL DESCRIPTION: We design a software resolving the following inverse problem: define the shape of a mirror which reflects the light from a source to a defined target, distribution and support of densities being prescribed. Classical applications include the conception of solar oven, public lightning, car headlights... Mathematical modeling of this problem, related to the optimal transport theory, takes the form of a nonlinear Monge-Ampere type PDE. The numerical resolution of these models remained until recently a largely open problem. MOKABAJOUR project aims to develop, using algorithms invented especially at Inria and LJK, a reflector design software more efficient than geometrical methods used so far. The final step is to realize and physically test prototype reflectors.

- Participants: Boris Thibert, Jean-David Benamou and Quentin Mérigot
- Contact: Jean-David Benamou
- URL: https://project.inria.fr/mokabajour/

# 6. New Results

## 6.1. Rank optimality for the Burer-Monteiro factorization

*I. Waldspurger, A. Watersw*

In [39], Numerically solving a large scale semidefinite program, in full generality, is a challenge: The complexity of generic algorithms blows up quickly with the size of the unknown matrix. Fortunately, in many situations, the solution of the program has low rank, and this can be exploited to achieve algorithmic speedups. The most classical way to do this is the Burer-Monteiro factorization, introduced in [77]. It consists in writing the unknown matrix as the product of low-rank factors, and optimizing the factors instead of the matrix itself. The first theoretical guarantees for this method appeared in [69], where it was shown that this strategy almost always succeeds when the size of the factors is of the order of the square root of the full matrix. In our article, we show that, up to a marginal improvement, this result is optimal: Contrarily to what numerical experiments might suggest, there exist situations where the method fails if the size of the factors is chosen smaller.

## 6.2. Representer theorems in variational problems

*C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. De Gournay, P. Weiss*

In [29], we have established a general principle which states that regularizing an inverse problem with a convex function yields solutions which are convex combinations of a small number of *atoms*. These atoms are identified with the extreme points and elements of the extreme rays of the regularizer level sets. An extension to a broader class of quasi-convex regularizers is also discussed. As a side result, we characterize the minimizers of the total gradient variation, describing the solutions of total variation problem as a superposition of indicator functions of simply connected sets. That result provides an explanation of the so-called *staircasing* phenomenon.

## 6.3. The Sliding Frank-Wolfe algorithm for Super-resolution Microscopy Imaging

*Q. Denoyelle, V. Duval, G. Peyré, E. Soubies*

In [32], we have studied the theoretical and numerical performance of the Sliding Frank-Wolfe, a novel optimization algorithm to solve the BLASSO sparse spikes super-resolution problem. The BLASSO is a continuous (*i.e.* off-the-grid or grid-less) counterpart to the well-known $\ell^1$ sparse regularisation method (also known as LASSO or Basis Pursuit). Our algorithm is a variation on the classical Frank-Wolfe (also known as conditional gradient) which follows a recent trend of interleaving convex optimization updates (corresponding to adding new spikes) with non-convex optimization steps (corresponding to moving the spikes). Our main theoretical result is that this algorithm terminates in a finite number of steps under a mild non-degeneracy hypothesis. We then target applications of this method to several instances of single molecule fluorescence imaging modalities, among which certain approaches rely heavily on the inversion of a Laplace transform. Our second theoretical contribution is the proof of the exact support recovery property of the BLASSO to invert the 1-D Laplace transform in the case of positive spikes. On the numerical side, we conclude this paper with an extensive study of the practical performance of the Sliding Frank-Wolfe on different instantiations of single molecule fluorescence imaging, including convolutive and non-convolutive (Laplace-like) operators. This shows the versatility and superiority of this method with respect to alternative sparse recovery techniques.

## 6.4. Approximation of variational problems with a convexity constraint by PDEs of Abreu type

*G. Carlier, T. Radice*

In [31], motivated by some variational problems subject to a convexity constraint, we consider an approximation using the logarithm of the Hessian determinant as a barrier for the constraint. We show that the minimizer of this penalization can be approached by solving a second boundary value problem for Abreu's equation which is a well-posed nonlinear fourth-order elliptic problem. More interestingly, a similar approximation result holds for the initial constrained variational problem.

## 6.5. Variational methods for tomographic reconstruction with few views

*M. Bergounioux, I. Abraham, R. Abraham, G. Carlier, E. Le Pennec, E. Trélat*

In [16], we deal with a severe ill posed problem, namely the reconstruction process of an image during tomography acquisition with (very) few views. We present different methods that we investigated during the past decade. They are based on variational analysis. This is a survey paper and we refer to the quoted papers for more details.

## 6.6. A differential approach to the multi-marginal Schrödinger system

*G. Carlier, M. Laborde*

In [30], we develop an elementary and self-contained differential approach, in an $L^\infty$ setting, for well-posedness (existence, uniqueness and smooth dependence with respect to the data) for the multi-marginal Schrödinger system which arises in the entropic regularization of optimal transport problems.

## 6.7. Minimal convex extensions and finite difference discretization of the quadratic Monge-Kantorovich problem

*J-D. Benamou, V. Duval*

In [15] we present an adaptation of the MA-LBR scheme to the Monge-Ampr`e equation with second boundary value condition, provided the target is a convex set. This yields a fast adaptive method to numerically solve the Optimal Transport problem between two absolutely continuous measures, the second of which has convex support. The proposed numerical method actually captures a specific Brenier solution which is minimal in some sense. We prove the convergence of the method as the grid stepsize vanishes and we show with numerical experiments that it is able to reproduce subtle properties of the Optimal Transport problem.

## 6.8. Second order models for optimal transport and cubic splines on the Wasserstein space

*J-D. Benamou, T. O. Gallouët, F-X. Vialard*

On the space of probability densities, we extend in [28] the Wasserstein geodesics to the case of higher-order interpolation such as cubic spline interpolation. After presenting the natural extension of cubic splines to the Wasserstein space, we propose a simpler approach based on the relaxation of the variational problem on the path space. We explore two different numerical approaches, one based on multi-marginal optimal transport and entropic regularization and the other based on semi-discrete optimal transport.

## 6.9. An entropy minimization approach to second-order variational mean-field games

*J-D. Benamou, G. Carlier, S. Di Marino, L. Nenna*

In [26] we propose a new viewpoint on variational mean-field games with diffusion and quadratic Hamiltonian. We show the equivalence of such mean-field games with a relative entropy minimization at the level of probabilities on curves. We also address the time-discretization of such problems, establish Gamma-Convergence results as the time step vanishes and propose an efficient algorithm relying on this entropic interpretation as well as on the Sinkhorn scaling algorithm.

## 6.10. Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm

*J-D. Benamou, G. Carlier, L. Nenna*

Starting from Brenier's relaxed formulation of the incompress-ible Euler equation in terms of geodesics in the group of measure-preserving diffeomorphisms, we propose in [27] a numerical method based on Sinkhorn's algorithm for the entropic regularization of optimal transport. We also make a detailed comparison of this entropic regular-ization with the so-called Bredinger entropic interpolation problem (see [1]). Numerical results in dimension one and two illustrate the feasibility of the method.

## 6.11. Testing Gaussian Process with Applications to Super-Resolution

*J.-M. Azaïs, Y. De Castro, S. Mourareau*

In [13], we introduce exact testing procedures on the mean of a Gaussian process $X$ derived from the outcomes of $\ell_1$-minimization over the space of complex valued measures. The process $X$ can be thought as the sum of two terms: first, the convolution between some kernel and a target atomic measure (mean of the process); second, a random perturbation by an additive centered Gaussian process. The first testing procedure considered is based on a dense sequence of grids on the index set of $X$ and we establish that it converges (as the grid step tends to zero) to a randomized testing procedure: the decision of the test depends on the observation $X$ and also on an independent random variable. The second testing procedure is based on the maxima and the Hessian of $X$ in a grid-less manner. We show that both testing procedures can be performed when the variance is unknown (and the correlation function of $X$ is known). These testing procedures can be used for the problem of deconvolution over the space of complex valued measures, and applications in frame of the Super-Resolution theory are presented. As a byproduct, numerical investigations may demonstrate that our grid-less method is more powerful (it detects sparse alternatives) than tests based on very thin grids.

## 6.12. Approximate Optimal Designs for Multivariate Polynomial Regression

*Y. De Castro, F. Gamboa, D. Henrion, R. Hess, J.-B Lasserre*

In [19], we introduce a new approach aiming at computing approximate optimal designs for multivariate polynomial regressions on compact (semi-algebraic) design spaces. We use the moment-sum-of-squares hierarchy of semidefinite programming problems to solve numerically the approximate optimal design problem. The geometry of the design is recovered via semidefinite programming duality theory. This article shows that the hierarchy converges to the approximate optimal design as the order of the hierarchy increases. Furthermore, we provide a dual certificate ensuring finite convergence of the hierarchy and showing that the approximate optimal design can be computed numerically with our method. As a byproduct, we revisit the equivalence theorem of the experimental design theory: it is linked to the Christoffel polynomial and it characterizes finite convergence of the moment-sum-of-square hierarchies.

## 6.13. Simulation of multiphase porous media flows with minimizing movement and finite volume schemes

*C .Cancès, T. O. Gallouët, M. Laborde, L. Monsaingeon*

In [17]: the Wasserstein gradient flow structure of the PDE system governing multiphase flows in porous media was recently highlighted in [85]. The model can thus be approximated by means of the minimizing movement (or JKO) scheme. We solve the JKO scheme using the ALG2-JKO scheme proposed in [55]. The numerical results are compared to a classical upstream mobility Finite Volume scheme, for which strong stability properties can be established.

## 6.14. An unbalanced optimal transport splitting scheme for general advection-reaction-diffusion problems

*T. O. Gallouët, M. Laborde, L. Monsaingeon*

In [21] the authors show that unbalanced optimal transport provides a convenient framework to handle reaction and diffusion processes in a unified metric framework. We use a constructive method, alternating minimizing movements for the Wasserstein distance and for the Fisher-Rao distance, and prove existence of weak solutions for general scalar reaction-diffusion-advection equations. We extend the approach to systems of multiple interacting species, and also consider an application to a very degenerate diffusion problem involving a Gamma-limit. Moreover, some numerical simulations are included.

## 6.15. Generalized compressible fluid flows and solutions of the Camassa-Holm variational model

*T. O. Gallouët, A. Natale, F-X. Vialard*

In [35] : The Camassa-Holm equation on a domain M $\in \mathbb{R}^d$ , in one of its possible multi-dimensional generalizations, describes geodesics on the group of diffeomorphisms with respect to the H(div) metric. It has been recently reformulated as a geodesic equation for the L2 metric on a subgroup of the diffeomorphism group of the cone over M. We use such an interpretation to construct an analogue of Brenier's generalized incompressible Euler flows for the Camassa-Holm equation. This involves describing the fluid motion using probability measures on the space of paths on the cone, so that particles are allowed to split and cross. Differently from Brenier's model, however, we are also able to account for compressibility by employing an explicit probabilistic representation of the Jacobian of the flow map. We formulate the boundary value problem associated to the Camassa-Holm equation using such generalized flows. We prove existence of solutions and that, for short times, smooth solutions of the Camassa-Holm equations are the unique solutions of our model. We propose a numerical scheme to construct generalized solutions on the cone and present some numerical results illustrating the relation between the generalized Camassa-Holm and incompressible Euler solutions.

## 6.16. The Camassa-Holm equation as an incompressible Euler equation: a geometric point of view

*T. O. Gallouët, F-X. Vialard*

In [23]: The group of diffeomorphisms of a compact manifold endowed with the $L^2$ metric acting on the space of probability densities gives a unifying framework for the incompressible Euler equation and the theory of optimal mass transport. Recently, several authors have extended optimal transport to the space of positive Radon measures where the Wasserstein-Fisher-Rao distance is a natural extension of the classical $L^2$-Wasserstein distance. In this paper, we show a similar relation between this unbalanced optimal transport problem and the Hdiv right-invariant metric on the group of diffeomorphisms, which corresponds to the Camassa-Holm (CH) equation in one dimension. On the optimal transport side, we prove a polar factorization theorem on the automorphism group of half-densities. Geometrically, our point of view provides an isometric embedding of the group of diffeomorphisms endowed with this right-invariant metric in the automorphisms group of the fiber bundle of half densities endowed with an $L^2$ type of cone metric. This leads to a new formulation of the (generalized) CH equation as a geodesic equation on an isotropy subgroup of this automorphisms group; On $S_1$, solutions to the standard CH thus give particular solutions of the incompressible Euler equation on a group of homeomorphisms of $\mathbb{R}^2$ which preserve a radial density that has a singularity at 0. An other application consists in proving that smooth solutions of the Euler-Arnold equation for the Hdiv right-invariant metric are length minimizing geodesics for sufficiently short times.

## 6.17. Variational Second-Order Interpolation on the Group of Diffeomorphisms with a Right-Invariant Metric

*F-X. Vialard*

In [38] we propose a variational framework in which the minimization of the acceleration on the group of diffeomorphisms endowed with a right-invariant metric is well-posed. It relies on constraining the acceleration to belong to a Sobolev space of higher-order than the order of the metric in order to gain compactness. It provides the theoretical guarantee of existence of minimizers which is compulsory for numerical simulations.

## 6.18. Interpolating between Optimal Transport and MMD using Sinkhorn Divergences

*J. Feydy, T. Séjourné, F.X. Vialard, S-I. Amari, A. Trouvé, G. Peyré*

In [33]: Comparing probability distributions is a fundamental problem in data sciences. Simple norms and divergences such as the total variation and the relative entropy only compare densities in a point-wise manner and fail to capture the geometric nature of the problem. In sharp contrast, Maximum Mean Discrepancies (MMD) and Optimal Transport distances (OT) are two classes of distances between measures that take into account the geometry of the underlying space and metrize the convergence in law. This paper studies the Sinkhorn divergences, a family of geometric divergences that interpolates between MMD and OT. Relying on a new notion of geometric entropy, we provide theoretical guarantees for these divergences: positivity, convexity and metrization of the convergence in law. On the practical side, we detail a numerical scheme that enables the large scale application of these divergences for machine learning: on the GPU, gradients of the Sinkhorn loss can be computed for batches of a million samples.

# 7. Partnerships and Cooperations

## 7.1. National Initiatives

### *7.1.1. ANR*

J-D. Benamou and G. Carlier are members of the ANR MFG (ANR-16-CE40-0015-01). Scientific topics of the project: Mean field analysis Analysis of the MFG systems and of the Master equation Numerical analysis Models and applications

J-D. Benamou G. Carlier F-X. Vialard and T. O. Gallouët are members of ANR MAGA (ANR-13-JS01-0007-01). The Monge-Ampère equation is a fully nonlinear elliptic equation, which plays a central role in geometry and in the theory of optimal transport. However, the singular and non-linear nature of the equation is a serious obstruction to its efficient numerical resolution. The first aim of the MAGA project is to study and to implement discretizations of optimal transport and Monge-Ampère equations which rely on tools from computational geometry (Laguerre diagrams). In a second step, these solvers will be applied to concrete problems from various fields involving optimal transport or Monge-Ampère equations such as computational physics: early universe reconstruction problem, congestion/incompressibility constraints economics: principal agent problems, geometry: variational problems over convex bodies, reflector and refractor design for non-imaging optics

T. O. Gallouët is member of the ANR GEOPOR (JCJC of C. Cancès) Scientific topic: geometrical approach, based on Wasserstein gradient flow, for multiphase flows in porous media. Theory and Numerics.

T. O. Gallouët is member of the ANR MESA (JCJC of M. Fathi) Scientific topic: Stein methods.

## 7.2. European Initiatives

### *7.2.1. FP7 & H2020 Projects*

J-D. Benamou and G. Rukhaia are members of the ROMSOC ITN.

## 7.3. International Research Visitors

### *7.3.1. Visits of International Scientists*

- Shuangjian Zhang, (PostDoc), Université de Toronto, June-August 2018.
- Clarice Poon, Imperial College London, January 2018
- Teresa Radice, Université de Naples, many short stays.

### *7.3.2. Visits to International Teams*

*7.3.2.1. Research Stays Abroad*

- P. Pegon was invited for 10 days to Penn State College by Alberto Bressan in order to start a collaboration on the theory of ramified transport and applications to biology, and to give lectures (2) in the seminar series on Computational and Applied Mathematics.
- G. Carlier was a John von Neumann invited Professor at TUM (Munich) in 2018.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. Member of the Organizing Committees*

- G. Carlier was in the organizing committee of the conference *Des mathématiques de la décision aux jeux à champ moyen* held in Honor of Jean-Michel Lasry at Dauphine in July 2018.
- T. O. Gallouët was in the organizing committee of the international conference *Modern mathematical methods for Data analysis* held at Liège (Belgium) in June 2018.
- F-X. Vialard was in the organizing committee of the December 2018 Banff, applied optimal transport, stochastic geometric mechanic and shapes conference.

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Reviewer*

V. Duval has reviewed contributions to the LVA/ICA conference.

### *8.1.3. Journal*

*8.1.3.1. Member of the Editorial Boards*

G. Carlier is in the board of Journal de l'Ecole Polytechnique, Applied Mathematics and Optimization, Mathematics and Financial Economics and Journal of Dynamic Games (starting end of 2018).

*8.1.3.2. Reviewer - Reviewing Activities*

- V. Duval has reviewed papers for the following journals: *ESAIM COCV*, *IEEE Trans. on Signal Processing*, *Information and Inference: a Journal of the IMA*, *Math. Stat. Learning*, *SIIMS*, *Inverse Problems on Imaging*.
- J-D. Benamou has reviewed papers for the following journals: *SINUM*, *Numerische Math.*, ...
- T. O. Gallouët has reviewed papers for ARMA, ...
- P. Pegon has reviewed articles for Journal de Mathématiques Pures et Appliquées and Journal of Functional Analysis.
- F-X. Vialard has reviewed papers for Numerische Math., Siam IS, IEEE TMI, Siaga, P AMS.

### *8.1.4. Invited Talks*

- I. Waldspurger gave talks at the Séminaire de mathématiques appliquées du collège de France, at the Mathematical Image Analysis conference (Berlin), at the Colloquium in Applied and Computational Mathematics of ETH Zurich, at the 7th international conference on computational harmonic analysis (Nashville) and at the SIAM conference on imaging science (Bologne).
- V. Duval has given invited talks at the MAGA workshop (Jan.), Journée Parcimonie Bordeaux of the GdR MIA (May), GdT StatNum at CEREMADE (May), SIAM Imaging Conference (June), ISMP Conference (July), Journées MAS (Aug.).
- J-D. Benamou has given invited talks at : Algebraic and geometric aspects of numerical methods for differential equations Workshop (Institut Mittag-Leffler, Djursholm), Workshop on Moving and Adaptative Meshes for Global Atmospheric Modelling (University of Reading), Gradient flows: challenges and new directions Workshop (ICMS, Edinburgh), Workshop on Monge-Ampère numerical resolution methods (TU Eindhoven), Shape Analysis, Stochastic Geometric Mechanics and Applied Optimal Transport (BIRS, Banff), MIGSAA Mini-Symposium in Optimal Transport and its Applications (University of Edimburgh).
- G. Carlier gave talks at TUM (Munich), workshop Optimisation and Machine Learning in Economics at UCL (Londres), workshop Entropies, the Geometry of Nonlinear Flows, and their Applications (Banff, Canada), SFB Colloquium, TUM, Munich, Workshop PDEs and optimal Transport (Essaouira), PDE seminar, Imperal College (Londres), Optimal Transport and Applications (Pise), ERC Readi closing conference (EHESS Paris).
- Y. De Castro has given invited talks at Société Mathématiques de France national conference, Ecole Normale Supérieure de Lyon, Oxford numerical analysis seminar (invited by Pr Tanner), Ecole des Ponts ParisTech, and Labex Bezout day.
- F-X. Vialard gave talks at siam Imaging Sciences in bologna june 2018, septembre 2018 Lisbonne workshop on optimal transport, geometric mechanics, october 2018, Labex Bézout, université paris-est and wasinvited to Mittag-Leffler institute for a conference on optimal transport, geometry and algebra.
- T. O. Gallouët gave talks at UPMC, Paris, Séminaire du Laboratoire Jacques-Louis Lions, ANR MAGA meeting, Paris and Workshop on New Developments in PDEs and Related Topics, Essaouira, Maroc.
- P. Pegon was invited for 10 days to Penn State College by Alberto Bressan in order to start a collaboration on the theory of ramified transport and applications to biology, and to give lectures (2) in the seminar series on Computational and Applied Mathematics.
- Andrea Natale gave the following talks:
  - Generalized H(div) geodesics and solutions of the Camassa-Holm equation. *BIRS Workshop: "Shape Analysis, Stochastic Geometric Mechanics and Applied Optimal Transport"*, 12/2018, Banff (Canada).

- An optimal transport approach for the Camassa-Holm variational model. *Canadian Mathematical Society Winter Meeting*, 12/2018, Vancouver (Canada).
- Generalized H(div) geodesics and solutions of the Camassa-Holm equation. *Oberwolfach Seminar: "Optimal Transport Theory and Hydrodynamics (from Euler to Monge and vice versa)"*, 10/2018, Oberwolfach (Germany).
- Generalized H(div) geodesics and solutions of the Camassa-Holm equation. *MokaMeeting*, 12/2018, Inria Paris (France).

### *8.1.5. Research Administration*

J-D. Benamou is a member of the *Commision de restauration locale*.

J-D. Benamou is the *Commission Bureau* referent for the 4th floor of building A.

J-D. Benamou is a member of PSL *Conseil Académique*.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Licence: I. Waldspurger, Analyse 1, 92 heures d'équivalent TD, L1, Université Paris Dauphine, France

Master: V. Duval, project supervision (M2), INSA de Rouen, France

Licence: T. O. Gallouët, Optimisation (TD, L3, Orsay), EDP (TD, 2 ème anné, ENS).

Master, Licence: G. Carlier taught algebra (Licence, Dauphine, 91h), transport and variational methods in economics (Dauphine, M2, 18h) and Convex duality and applications in mass transport and the calculus of variations in Munich (18h).

Licence: Y. De Castro taught optimization to first year engineering school student (10h).

Licence: A. Natale taught (TD) for the course "Calcul Différentiel et Optimisation" (Instructor: Emeric Bouin), 09-12/2018, Université Paris Dauphine, Paris (France).

Licence, Master: P. Pegon taught a crash course in analysis (M1, Dauphine, 12h), and gave exercise sessions in Measure theory and probability (L3, Dauphine, 39h) and on Functional Analysis and PDEs (M1, Dauphine, 19.5h).

### *8.2.2. Supervision*

- PhD : Quentin Denoyelle, *Theoretical and Numerical Analysis of Super-Resolution without Grid*, defended on 09/07/2018, G. Peyré and V. Duval.
- PhD in progress : Gabriele Todeschi, Optimal transport and finite volumes, 01/10/2018, T. O. Gallouët.
- PhD in progress : Miao Yu , *Optimal Transport distances for Full Waveform Inversion*, 01/10/2016, J-D. Benamou.
- PhD in progress: Paul Catala, *Low-rank Approaches for Off-the-grid Superresolution*, 01/10/2016, G. Peyré and V. Duval.
- PhD in progress : Lucas Martinet , *Decomposed and Parallel Sinkhorn Algorithm for Optimal Transport*, 01/10/2017, J-D. Benamou.
- PhD in progress : Giorgi Rukhaia , *On the application of Sinkhorn methods in Freeform Optics*, 01/05/2018, J-D. Benamou.
- PhD in progress : Ernesto Araya , *Random Graphs*, Y. De Castro.

### *8.2.3. Juries*

G. Carlier was in the HDR committee of Nicolas Juillet and in the Ph.D committees of Michael Orieux, Hadrien de March and Thomas Dumas

## 8.3. Popularization

### *8.3.1. Internal or external Inria responsibilities*

V. Duval was a member of the *Commission d'évaluation scientifique* (CES) of the CRI Paris in 2018, and has been a member of the *Comité de suivi doctoral* since September 2018.

J-D. Benamou was a member of the *CR2 Recruiting Commission* of the CRI Paris in 2018.

### *8.3.2. Education*

I. Waldspurger has given a talk at the Mathematic Park seminar (for L1 and L2 students), on the minimization of convex functions with Lipschitz gradient.

### *8.3.3. Interventions*

Participation à l'accueil de 15 collégiens dans le cadre de leur stage d'observation (3 heures dans l'équipe MOKAPLAN supervisé par Lucas Martinet).

# 9. Bibliography

## Major publications by the team in recent years

[1] M. AGUEH, G. CARLIER.*Barycenters in the Wasserstein space*, in "SIAM J. Math. Anal.", 2011, vol. 43, n$^{o}$ 2, p. 904–924

[2] J.-D. BENAMOU, Y. BRENIER.*A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, in "Numer. Math.", 2000, vol. 84, n$^{o}$ 3, p. 375–393, http://dx.doi.org/10.1007/s002110050002

[3] J.-D. BENAMOU, G. CARLIER, M. CUTURI, L. NENNA, G. PEYRÉ.*Iterative Bregman Projections for Regularized Transportation Problems*, in "SIAM Journal on Scientific Computing", 2015, vol. 37, n$^{o}$ 2, p. A1111-A1138 [*DOI :* 10.1137/141000439], http://hal.archives-ouvertes.fr/hal-01096124

[4] J.-D. BENAMOU, F. COLLINO, J.-M. MIREBEAU.*Monotone and Consistent discretization of the Monge-Ampere operator*, September 2014, pubished in MAth of Comp, https://hal.archives-ouvertes.fr/hal-01067540

[5] M. BRUVERIS, F.-X. VIALARD.*On Completeness of Groups of Diffeomorphisms*, in "ArXiv e-prints", March 2014

[6] V. DUVAL, G. PEYRÉ.*Exact Support Recovery for Sparse Spikes Deconvolution*, in "Foundations of Computational Mathematics", 2014, p. 1-41, http://dx.doi.org/10.1007/s10208-014-9228-6

[7] F. GAY-BALMAZ, D. D. HOLM, D. M. MEIER, T. S. RATIU, F.-X. VIALARD.*Invariant Higher-Order Variational Problems*, in "Communications in Mathematical Physics", January 2012, vol. 309, p. 413-458, http://dx.doi.org/10.1007/s00220-011-1313-y

[8] P. MACHADO MANHÃES DE CASTRO, Q. MÉRIGOT, B. THIBERT.*Intersection of paraboloids and application to Minkowski-type problems*, in "Numerische Mathematik", November 2015 [*DOI :* 10.1007/s00211-015-0780-z], https://hal.archives-ouvertes.fr/hal-00952720

[9] Q. MÉRIGOT.*A multiscale approach to optimal transport*, in "Computer Graphics Forum", 2011, vol. 30, n$^{o}$ 5, p. 1583–1592

[10] I. WALDSPURGER, A. WATERS.*Rank optimality for the Burer-Monteiro factorization*, December 2018, preprint, https://hal.archives-ouvertes.fr/hal-01958814

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[11] Q. DENOYELLE.*Theoretical and Numerical Analysis of Super-Resolution Without Grid*, PSL Research University, July 2018, https://tel.archives-ouvertes.fr/tel-02002504

## Articles in International Peer-Reviewed Journal

[12] J.-M. AZAÏS, Y. DE CASTRO, Y. GOUDE, G. HÉBRAIL, J. MEI.*Nonnegative matrix factorization with side information for time series recovery and prediction*, in "IEEE Transactions on Knowledge and Data Engineering", May 2018, https://arxiv.org/abs/1709.06320 , https://hal.inria.fr/hal-01686429

[13] J.-M. AZAÏS, Y. DE CASTRO, S. MOURAREAU.*Testing Gaussian Process with Applications to Super-Resolution*, in "Applied and Computational Harmonic Analysis", July 2018, https://arxiv.org/abs/1706.00679 - Final versio, Python code and Jupyter notebook available at https://github.com/ydecastro/super-resolution-testing, https://hal.inria.fr/hal-01686434

[14] J.-D. BENAMOU, G. CARLIER, M. LABORDE.*An augmented Lagrangian approach to Wasserstein gradient flows and applications*, in "ESAIM: Proceedings and Surveys", August 2019, https://hal.archives-ouvertes.fr/hal-01245184

[15] J.-D. BENAMOU, V. DUVAL.*Minimal convex extensions and finite difference discretization of the quadratic Monge-Kantorovich problem*, in "European Journal of Applied Mathematics", 2019, https://arxiv.org/abs/1710.05594 [*DOI :* 10.1017/S0956792518000451], https://hal.inria.fr/hal-01616842

[16] M. BERGOUNIOUX, I. ABRAHAM, R. ABRAHAM, G. CARLIER, E. LE PENNEC, E. TRÉLAT.*Variational methods for tomographic reconstruction with few views*, in "Milan Journal of Mathematics", 2018, vol. 86, n$^{o}$ 2, p. 157–200, https://hal.archives-ouvertes.fr/hal-01817172

[17] 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

[18] S. DALLAPORTA, Y. DE CASTRO.*Sparse Recovery from Extreme Eigenvalues Deviation Inequalities*, in "ESAIM: Probability and Statistics", 2019, https://arxiv.org/abs/1604.01171 - 33 pages, 1 figure, https://hal.archives-ouvertes.fr/hal-01309439

[19] Y. DE CASTRO, F. GAMBOA, D. HENRION, R. HESS, J.-B. LASSERRE.*Approximate Optimal Designs for Multivariate Polynomial Regression*, in "Annals of Statistics", January 2019, vol. 47, n$^{o}$ 1, p. 127-155, https://hal.laas.fr/hal-01483490

[20] S. DI MARINO, J. LOUET.*The entropic regularization of the Monge problem on the real line*, in "SIAM Journal on Mathematical Analysis", July 2018, vol. 50, n$^{o}$ 4, p. 3451 - 3477, https://arxiv.org/abs/1703.10457 [*DOI :* 10.1137/17M1123523], https://hal.archives-ouvertes.fr/hal-01498732

[21] T. GALLOUËT, M. LABORDE, L. MONSAINGEON.*An unbalanced optimal transport splitting scheme for general advection-reaction-diffusion problems*, in "ESAIM: Control, Optimisation and Calculus of Variations", 2019, https://arxiv.org/abs/1704.04541 , https://hal.archives-ouvertes.fr/hal-01508911

[22] T. O. GALLOUËT, Q. MÉRIGOT.*A Lagrangian Scheme à la Brenier for the Incompressible Euler Equations*, in "Foundations of Computational Mathematics", 2018, https://arxiv.org/abs/1605.00568 [*DOI :* 10.1007/s10208-017-9355-y], https://hal.archives-ouvertes.fr/hal-01425826

[23] T. GALLOUËT, F.-X. VIALARD.*The Camassa-Holm equation as an incompressible Euler equation: a geometric point of view*, in "Journal of Differential Equations", April 2018, https://arxiv.org/abs/1609.04006 , https://hal.archives-ouvertes.fr/hal-01363647

### Conferences without Proceedings

[24] J.-B. COURBOT, E. MONFRINI, V. MAZET, C. COLLET.*Triplet markov trees for image segmentation*, in "2018 IEEE Workshop on Statistical Signal Processing (SSP 2018)", Fribourg-en-Brisgau, Germany, June 2018, https://hal.archives-ouvertes.fr/hal-01815562

[25] J. M. FADILI, G. GARRIGOS, J. MALICK, G. PEYRÉ.*Model Consistency for Learning with Mirror-Stratifiable Regularizers*, in "International Conference on Artificial Intelligence and Statistics (AISTATS)", Naha, Japan, April 2019, https://hal.archives-ouvertes.fr/hal-01988309

### Other Publications

[26] J.-D. BENAMOU, G. CARLIER, S. DI MARINO, L. NENNA.*An entropy minimization approach to second-order variational mean-field games*, August 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01848370

[27] J.-D. BENAMOU, G. CARLIER, L. NENNA.*Generalized incompressible flows, multi-marginal transport and Sinkhorn algorithm*, March 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01621311

[28] J.-D. BENAMOU, T. GALLOUËT, F.-X. VIALARD.*Second order models for optimal transport and cubic splines on the Wasserstein space*, January 2018, https://arxiv.org/abs/1801.04144 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01682107

[29] C. BOYER, A. CHAMBOLLE, Y. DE CASTRO, V. DUVAL, F. DE GOURNAY, P. WEISS.*On Representer Theorems and Convex Regularization*, November 2018, https://arxiv.org/abs/1806.09810 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01823135

[30] G. CARLIER, M. LABORDE.*A differential approach to the multi-marginal Schrödinger system*, November 2018, https://arxiv.org/abs/1811.05207 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01918578

[31] G. CARLIER, T. RADICE.*Approximation of variational problems with a convexity constraint by PDEs of Abreu type*, May 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01802925

[32] Q. DENOYELLE, V. DUVAL, G. PEYRÉ, E. SOUBIES.*The Sliding Frank-Wolfe Algorithm and its Application to Super-Resolution Microscopy*, November 2018, https://arxiv.org/abs/1811.06416 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01921604

[33] J. FEYDY, T. SÉJOURNÉ, F.-X. VIALARD, S.-I. AMARI, A. TROUVÉ, G. PEYRÉ.*Interpolating between Optimal Transport and MMD using Sinkhorn Divergences*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01898858

[34] T. GALLOUËT, G. MIJOULE, Y. SWAN.*Regularity of solutions of the Stein equation and rates in the multivariate central limit theorem*, May 2018, https://arxiv.org/abs/1805.01720 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01785397

[35] T. GALLOUËT, A. NATALE, F.-X. VIALARD.*Generalized compressible fluid flows and solutions of the Camassa-Holm variational model* , September 2018, https://arxiv.org/abs/1806.10825 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01815531

[36] M. MASOERO.*On the long time convergence of potential MFG*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01850874

[37] A. NATALE, F.-X. VIALARD.*Embedding Camassa-Holm equations in incompressible Euler*, April 2018, https://arxiv.org/abs/1804.11080 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01781162

[38] F.-X. VIALARD.*Variational Second-Order Interpolation on the Group of Diffeomorphisms with a Right-Invariant Metric*, January 2018, https://arxiv.org/abs/1801.04146 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01682108

[39] I. WALDSPURGER, A. WATERS.*Rank optimality for the Burer-Monteiro factorization*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01958814

## References in notes

[40] I. ABRAHAM, R. ABRAHAM, M. BERGOUNIOUX, G. CARLIER.*Tomographic reconstruction from a few views: a multi-marginal optimal transport approach*, in "Preprint Hal-01065981", 2014

[41] Y. ACHDOU, V. PEREZ.*Iterative strategies for solving linearized discrete mean field games systems*, in "Netw. Heterog. Media", 2012, vol. 7, n$^{o}$ 2, p. 197–217, http://dx.doi.org/10.3934/nhm.2012.7.197

[42] M. AGUEH, G. CARLIER.*Barycenters in the Wasserstein space*, in "SIAM J. Math. Anal.", 2011, vol. 43, n$^{o}$ 2, p. 904–924

[43] F. ALTER, V. CASELLES, A. CHAMBOLLE.*Evolution of Convex Sets in the Plane by Minimizing the Total Variation Flow*, in "Interfaces and Free Boundaries", 2005, vol. 332, p. 329–366

[44] F. R. BACH.*Consistency of the Group Lasso and Multiple Kernel Learning*, in "J. Mach. Learn. Res.", June 2008, vol. 9, p. 1179–1225, http://dl.acm.org/citation.cfm?id=1390681.1390721

[45] F. R. BACH.*Consistency of Trace Norm Minimization*, in "J. Mach. Learn. Res.", June 2008, vol. 9, p. 1019–1048, http://dl.acm.org/citation.cfm?id=1390681.1390716

[46] H. H. BAUSCHKE, P. L. COMBETTES.*A Dykstra-like algorithm for two monotone operators*, in "Pacific Journal of Optimization", 2008, vol. 4, n$^{o}$ 3, p. 383–391

[47] M. F. BEG, M. I. MILLER, A. TROUVÉ, L. YOUNES.*Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms*, in "International Journal of Computer Vision", February 2005, vol. 61, n$^{o}$ 2, p. 139–157, http://dx.doi.org/10.1023/B:VISI.0000043755.93987.aa

[48] M. BEIGLBOCK, P. HENRY-LABORDÈRE, F. PENKNER.*Model-independent bounds for option prices mass transport approach*, in "Finance and Stochastics", 2013, vol. 17, n$^{o}$ 3, p. 477-501, http://dx.doi.org/10.1007/s00780-013-0205-8

[49] G. BELLETTINI, V. CASELLES, M. NOVAGA.*The Total Variation Flow in $R^N$*, in "J. Differential Equations", 2002, vol. 184, n$^{o}$ 2, p. 475–525

[50] J.-D. BENAMOU, Y. BRENIER.*A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem*, in "Numer. Math.", 2000, vol. 84, n$^{o}$ 3, p. 375–393, http://dx.doi.org/10.1007/s002110050002

[51] J.-D. BENAMOU, Y. BRENIER.*Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampère/transport problem*, in "SIAM J. Appl. Math.", 1998, vol. 58, n$^{o}$ 5, p. 1450–1461

[52] J.-D. BENAMOU, G. CARLIER.*Augmented Lagrangian algorithms for variational problems with divergence constraints*, in "JOTA", 2015

[53] J.-D. BENAMOU, G. CARLIER, N. BONNE.*An Augmented Lagrangian Numerical approach to solving Mean-Fields Games*, Inria, December 2013, 30, http://hal.inria.fr/hal-00922349

[54] J.-D. BENAMOU, G. CARLIER, M. CUTURI, L. NENNA, G. PEYRÉ.*Iterative Bregman Projections for Regularized Transportation Problems*, in "SIAM J. Sci. Comp.", 2015, to appear

[55] J.-D. BENAMOU, G. CARLIER, M. LABORDE.*An augmented Lagrangian approach to Wasserstein gradient flows and applications*, in "ESAIM: Proceedings and Surveys", August 2019, https://hal.archives-ouvertes.fr/hal-01245184

[56] J.-D. BENAMOU, G. CARLIER, Q. MÉRIGOT, É. OUDET.*Discretization of functionals involving the Monge-Ampère operator*, HAL, July 2014, https://hal.archives-ouvertes.fr/hal-01056452

[57] J.-D. BENAMOU, F. COLLINO, J.-M. MIREBEAU.*Monotone and Consistent discretization of the Monge-Ampère operator*, in "arXiv preprint arXiv:1409.6694", 2014, to appear in Math of Comp

[58] J.-D. BENAMOU, B. D. FROESE, A. OBERMAN.*Two numerical methods for the elliptic Monge-Ampère equation*, in "M2AN Math. Model. Numer. Anal.", 2010, vol. 44, n$^{o}$ 4, p. 737–758

[59] J.-D. BENAMOU, B. D. FROESE, A. OBERMAN.*Numerical solution of the optimal transportation problem using the Monge–Ampere equation*, in "Journal of Computational Physics", 2014, vol. 260, p. 107–126

[60] F. BENMANSOUR, G. CARLIER, G. PEYRÉ, F. SANTAMBROGIO.*Numerical approximation of continuous traffic congestion equilibria*, in "Netw. Heterog. Media", 2009, vol. 4, n$^{o}$ 3, p. 605–623

[61] M. BENNING, M. BURGER.*Ground states and singular vectors of convex variational regularization methods*, in "Meth. Appl. Analysis", 2013, vol. 20, p. 295–334

[62] B. BERKELS, A. EFFLAND, M. RUMPF.*Time discrete geodesic paths in the space of images*, in "Arxiv preprint", 2014

[63] J. BIGOT, T. KLEIN.*Consistent estimation of a population barycenter in the Wasserstein space*, in "Preprint arXiv:1212.2562", 2012

[64] A. BLANCHET, P. LAURENÇOT.*The parabolic-parabolic Keller-Segel system with critical diffusion as a gradient flow in $R^d$, $d \geq 3$*, in "Comm. Partial Differential Equations", 2013, vol. 38, n$^{o}$ 4, p. 658–686, http://dx.doi.org/10.1080/03605302.2012.757705

[65] J. BLEYER, G. CARLIER, V. DUVAL, J.-M. MIREBEAU, G. PEYRÉ.*A $\Gamma$ Convergence Result for the Upper Bound Limit Analysis of Plates*, in "ESAIM: Mathematical Modelling and Numerical Analysis", January 2016, vol. 50, n$^{o}$ 1, p. 215–235 [*DOI :* 10.1051/M2AN/2015040], https://www.esaim-m2an.org/articles/m2an/abs/2016/01/m2an141087/m2an141087.html

[66] N. BONNEEL, J. RABIN, G. PEYRÉ, H. PFISTER.*Sliced and Radon Wasserstein Barycenters of Measures*, in "Journal of Mathematical Imaging and Vision", 2015, vol. 51, n$^{o}$ 1, p. 22–45, http://hal.archives-ouvertes.fr/hal-00881872/

[67] U. BOSCAIN, R. CHERTOVSKIH, J.-P. GAUTHIER, D. PRANDI, A. REMIZOV.*Highly corrupted image inpainting through hypoelliptic diffusion*, Preprint CMAP, 2014, http://hal.archives-ouvertes.fr/hal-00842603/

[68] G. BOUCHITTÉ, G. BUTTAZZO.*Characterization of optimal shapes and masses through Monge-Kantorovich equation*, in "J. Eur. Math. Soc. (JEMS)", 2001, vol. 3, n$^{o}$ 2, p. 139–168, http://dx.doi.org/10.1007/s100970000027

[69] N. BOUMAL, V. VORONINSKI, A. S. BANDEIRA.*Deterministic guarantees for Burer-Monteiro factorizations of smooth semidefinite programs*, in "preprint", 2018, https://arxiv.org/abs/1804.02008

[70] L. BRASCO, G. CARLIER, F. SANTAMBROGIO.*Congested traffic dynamics, weak flows and very degenerate elliptic equations*, in "J. Math. Pures Appl. (9)", 2010, vol. 93, n$^{o}$ 6, p. 652–671

[71] L. M. BREGMAN.*The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, in "USSR computational mathematics and mathematical physics", 1967, vol. 7, n$^{o}$ 3, p. 200–217

[72] Y. BRENIER.*Generalized solutions and hydrostatic approximation of the Euler equations*, in "Phys. D", 2008, vol. 237, n$^{o}$ 14-17, p. 1982–1988, http://dx.doi.org/10.1016/j.physd.2008.02.026

[73] Y. BRENIER.*Décomposition polaire et réarrangement monotone des champs de vecteurs*, in "C. R. Acad. Sci. Paris Sér. I Math.", 1987, vol. 305, n$^{o}$ 19, p. 805–808

[74] Y. BRENIER.*Polar factorization and monotone rearrangement of vector-valued functions*, in "Comm. Pure Appl. Math.", 1991, vol. 44, n$^{o}$ 4, p. 375–417, http://dx.doi.org/10.1002/cpa.3160440402

[75] Y. BRENIER, U. FRISCH, M. HENON, G. LOEPER, S. MATARRESE, R. MOHAYAEE, A. SOBOLEVSKI.*Reconstruction of the early universe as a convex optimization problem*, in "Mon. Not. Roy. Astron. Soc.", 2003, vol. 346, p. 501–524, http://arxiv.org/pdf/astro-ph/0304214.pdf

[76] M. BRUVERIS, L. RISSER, F.-X. VIALARD.*Mixture of Kernels and Iterated Semidirect Product of Diffeomorphisms Groups*, in "Multiscale Modeling & Simulation", 2012, vol. 10, n$^{o}$ 4, p. 1344-1368

[77] S. BURER, R. D. C. MONTEIRO.*A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization*, in "Mathematical Programming", 2003, vol. 95, n$^{o}$ 2, p. 329-357

[78] M. BURGER, M. DI FRANCESCO, P. MARKOWICH, M. T. WOLFRAM.*Mean field games with nonlinear mobilities in pedestrian dynamics*, in "DCDS B", 2014, vol. 19

[79] M. BURGER, M. FRANEK, C.-B. SCHÖNLIEB.*Regularized regression and density estimation based on optimal transport*, in "Appl. Math. Res. Expr.", 2012, vol. 2, p. 209–253

[80] M. BURGER, S. OSHER.*A guide to the TV zoo*, in "Level-Set and PDE-based Reconstruction Methods, Springer", 2013

[81] G. BUTTAZZO, C. JIMENEZ, É. OUDET.*An optimization problem for mass transportation with congested dynamics*, in "SIAM J. Control Optim.", 2009, vol. 48, n$^{o}$ 3, p. 1961–1976

[82] H. BYRNE, D. DRASDO.*Individual-based and continuum models of growing cell populations: a comparison*, in "Journal of Mathematical Biology", 2009, vol. 58, n$^{o}$ 4-5, p. 657-687

[83] L. A. CAFFARELLI.*The regularity of mappings with a convex potential*, in "J. Amer. Math. Soc.", 1992, vol. 5, n$^{o}$ 1, p. 99–104, http://dx.doi.org/10.2307/2152752

[84] L. A. CAFFARELLI, S. A. KOCHENGIN, V. OLIKER.*On the numerical solution of the problem of reflector design with given far-field scattering data*, in "Monge Ampère equation: applications to geometry and optimization (Deerfield Beach, FL, 1997)", Providence, RI, Contemp. Math., Amer. Math. Soc., 1999, vol. 226, p. 13–32, http://dx.doi.org/10.1090/conm/226/03233

[85] C. CANCÈS, T. GALLOUËT, L. MONSAINGEON.*Incompressible immiscible multiphase flows in porous media: a variational approach*, in "Analysis & PDE", 2017, vol. 10, n$^{o}$ 8, p. 1845–1876 [*DOI :* 10.2140/APDE.2017.10.1845], https://hal.archives-ouvertes.fr/hal-01345438

[86] E. J. CANDÈS, C. FERNANDEZ-GRANDA.*Super-Resolution from Noisy Data*, in "Journal of Fourier Analysis and Applications", 2013, vol. 19, n$^{o}$ 6, p. 1229–1254

[87] E. J. CANDÈS, C. FERNANDEZ-GRANDA.*Towards a Mathematical Theory of Super-Resolution*, in "Communications on Pure and Applied Mathematics", 2014, vol. 67, n$^{o}$ 6, p. 906–956

[88] E. J. CANDÈS, M. WAKIN.*An Introduction to Compressive Sensing*, in "IEEE Signal Processing Magazine", 2008, vol. 25, n$^{o}$ 2, p. 21–30

[89] P. CARDALIAGUET, G. CARLIER, B. NAZARET.*Geodesics for a class of distances in the space of probability measures*, in "Calc. Var. Partial Differential Equations", 2013, vol. 48, n$^{o}$ 3-4, p. 395–420

[90] G. CARLIER.*A general existence result for the principal-agent problem with adverse selection*, in "J. Math. Econom.", 2001, vol. 35, n$^{o}$ 1, p. 129–150

[91] G. CARLIER, V. CHERNOZHUKOV, A. GALICHON.*Vector Quantile Regression*, Arxiv 1406.4643, 2014

[92] G. CARLIER, M. COMTE, I. IONESCU, G. PEYRÉ.*A Projection Approach to the Numerical Analysis of Limit Load Problems*, in "Mathematical Models and Methods in Applied Sciences", 2011, vol. 21, n$^{o}$ 6, p. 1291–1316 [*DOI :* DOI:10.1142/S0218202511005325], http://hal.archives-ouvertes.fr/hal-00450000/

[93] G. CARLIER, X. DUPUIS.*An iterated projection approach to variational problems under generalized convexity constraints and applications*, In preparation, 2015

[94] G. CARLIER, C. JIMENEZ, F. SANTAMBROGIO.*Optimal Transportation with Traffic Congestion and Wardrop Equilibria*, in "SIAM Journal on Control and Optimization", 2008, vol. 47, n$^{o}$ 3, p. 1330-1350

[95] G. CARLIER, T. LACHAND-ROBERT, B. MAURY.*A numerical approach to variational problems subject to convexity constraint*, in "Numer. Math.", 2001, vol. 88, n$^{o}$ 2, p. 299–318, http://dx.doi.org/10.1007/PL00005446

[96] G. CARLIER, A. OBERMAN, É. OUDET.*Numerical methods for matching for teams and Wasserstein barycenters*, in "M2AN", 2015, to appear

[97] J. A. CARRILLO, S. LISINI, E. MAININI.*Uniqueness for Keller-Segel-type chemotaxis models*, in "Discrete Contin. Dyn. Syst.", 2014, vol. 34, n$^{o}$ 4, p. 1319–1338, http://dx.doi.org/10.3934/dcds.2014.34.1319

[98] V. CASELLES, A. CHAMBOLLE, M. NOVAGA.*The discontinuity set of solutions of the TV denoising problem and some extensions*, in "Multiscale Modeling and Simulation", 2007, vol. 6, n$^{o}$ 3, p. 879–894

[99] C. CERITOGLU, E. AL..*Computational Analysis of LDDMM for Brain Mapping*, in "Frontiers in Neuroscience", 2013, vol. 7

[100] F. A. C. C. CHALUB, P. A. MARKOWICH, B. PERTHAME, C. SCHMEISER.*Kinetic models for chemotaxis and their drift-diffusion limits*, in "Monatsh. Math.", 2004, vol. 142, n$^{o}$ 1-2, p. 123–141, http://dx.doi.org/10.1007/s00605-004-0234-7

[101] A. CHAMBOLLE, T. POCK.*On the ergodic convergence rates of a first-order primal-dual algorithm*, in "Preprint OO/2014/09/4532", 2014

[102] G. CHARPIAT, G. NARDI, G. PEYRÉ, F.-X. VIALARD.*Finsler Steepest Descent with Applications to Piecewise-regular Curve Evolution*, Preprint hal-00849885, 2013, http://hal.archives-ouvertes.fr/hal-00849885/

[103] S. S. CHEN, D. L. DONOHO, M. A. SAUNDERS.*Atomic decomposition by basis pursuit*, in "SIAM journal on scientific computing", 1999, vol. 20, n$^{o}$ 1, p. 33–61

[104] P. CHONÉ, H. V. J. LE MEUR.*Non-convergence result for conformal approximation of variational problems subject to a convexity constraint*, in "Numer. Funct. Anal. Optim.", 2001, vol. 22, n$^{o}$ 5-6, p. 529–547, http://dx.doi.org/10.1081/NFA-100105306

[105] C. COTAR, G. FRIESECKE, C. KLUPPELBERG.*Density Functional Theory and Optimal Transportation with Coulomb Cost*, in "Communications on Pure and Applied Mathematics", 2013, vol. 66, n$^{o}$ 4, p. 548–599, http://dx.doi.org/10.1002/cpa.21437

[106] M. J. P. CULLEN, W. GANGBO, G. PISANTE.*The semigeostrophic equations discretized in reference and dual variables*, in "Arch. Ration. Mech. Anal.", 2007, vol. 185, n$^{o}$ 2, p. 341–363, http://dx.doi.org/10.1007/s00205-006-0040-6

[107] M. J. P. CULLEN, J. NORBURY, R. J. PURSER.*Generalised Lagrangian solutions for atmospheric and oceanic flows*, in "SIAM J. Appl. Math.", 1991, vol. 51, n$^{o}$ 1, p. 20–31

[108] M. CUTURI.*Sinkhorn Distances: Lightspeed Computation of Optimal Transport*, in "Proc. NIPS", C. J. C. BURGES, L. BOTTOU, Z. GHAHRAMANI, K. Q. WEINBERGER (editors), 2013, p. 2292–2300

[109] E. J. DEAN, R. GLOWINSKI.*Numerical methods for fully nonlinear elliptic equations of the Monge-Ampère type*, in "Comput. Methods Appl. Mech. Engrg.", 2006, vol. 195, n$^{o}$ 13-16, p. 1344–1386

[110] V. DUVAL, G. PEYRÉ.*Exact Support Recovery for Sparse Spikes Deconvolution*, in "Foundations of Computational Mathematics", 2014, p. 1-41, http://dx.doi.org/10.1007/s10208-014-9228-6

[111] V. DUVAL, G. PEYRÉ.*Sparse regularization on thin grids I: the L asso*, in "Inverse Problems", 2017, vol. 33, n$^{o}$ 5, 055008 [*DOI :* 10.1088/1361-6420/AA5E12], http://stacks.iop.org/0266-5611/33/i=5/a=055008

[112] J. FEHRENBACH, J.-M. MIREBEAU.*Sparse Non-negative Stencils for Anisotropic Diffusion*, in "Journal of Mathematical Imaging and Vision", 2014, vol. 49, n$^{o}$ 1, p. 123-147, http://dx.doi.org/10.1007/s10851-013-0446-3

[113] C. FERNANDEZ-GRANDA.*Support detection in super-resolution*, in "Proc. Proceedings of the 10th International Conference on Sampling Theory and Applications", 2013, p. 145–148

[114] A. FIGALLI, R. MCCANN, Y. KIM.*When is multi-dimensional screening a convex program?*, in "Journal of Economic Theory", 2011

[115] J.-B. FIOT, H. RAGUET, L. RISSER, L. D. COHEN, J. FRIPP, F.-X. VIALARD.*Longitudinal deformation models, spatial regularizations and learning strategies to quantify Alzheimer's disease progression*, in "NeuroImage: Clinical", 2014, vol. 4, n$^{o}$ 0, p. 718 - 729 [*DOI :* 10.1016/J.NICL.2014.02.002], http://www.sciencedirect.com/science/article/pii/S2213158214000205

[116] J.-B. FIOT, L. RISSER, L. D. COHEN, J. FRIPP, F.-X. VIALARD.*Local vs Global Descriptors of Hippocampus Shape Evolution for Alzheimer's Longitudinal Population Analysis*, in "Spatio-temporal Image Analysis for Longitudinal and Time-Series Image Data", Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, vol. 7570, p. 13-24, http://dx.doi.org/10.1007/978-3-642-33555-6_2

[117] U. FRISCH, S. MATARRESE, R. MOHAYAEE, A. SOBOLEVSKI.*Monge-Ampère-Kantorovitch (MAK) reconstruction of the eary universe*, in "Nature", 2002, vol. 417, n$^{o}$ 260

[118] B. D. FROESE, A. OBERMAN.*Convergent filtered schemes for the Monge-Ampère partial differential equation*, in "SIAM J. Numer. Anal.", 2013, vol. 51, n$^{o}$ 1, p. 423–444

[119] A. GALICHON, P. HENRY-LABORDÈRE, N. TOUZI.*A stochastic control approach to No-Arbitrage bounds given marginals, with an application to Loopback options*, in "submitted to Annals of Applied Probability", 2011

[120] W. GANGBO, R. MCCANN.*The geometry of optimal transportation*, in "Acta Math.", 1996, vol. 177, n$^{o}$ 2, p. 113–161, http://dx.doi.org/10.1007/BF02392620

[121] E. GHYS.*Gaspard Monge, Le mémoire sur les déblais et les remblais*, in "Image des mathématiques, CNRS", 2012, http://images.math.cnrs.fr/Gaspard-Monge,1094.html

[122] O. GUÉANT, J.-M. LASRY, P.-L. LIONS.*Mean field games and applications*, in "Paris-Princeton Lectures on Mathematical Finance 2010", Berlin, Lecture Notes in Math., Springer, 2011, vol. 2003, p. 205–266, http://dx.doi.org/10.1007/978-3-642-14660-2_3

[123] G. HERMAN.*Image reconstruction from projections: the fundamentals of computerized tomography*, Academic Press, 1980

[124] D. D. HOLM, J. T. RATNANATHER, A. TROUVÉ, L. YOUNES.*Soliton dynamics in computational anatomy*, in "NeuroImage", 2004, vol. 23, p. S170–S178

[125] B. J. HOSKINS.*The mathematical theory of frontogenesis*, in "Annual review of fluid mechanics, Vol. 14", Palo Alto, CA, Annual Reviews, 1982, p. 131–151

[126] R. JORDAN, D. KINDERLEHRER, F. OTTO.*The variational formulation of the Fokker-Planck equation*, in "SIAM J. Math. Anal.", 1998, vol. 29, n$^{o}$ 1, p. 1–17

[127] W. JÄGER, S. LUCKHAUS.*On explosions of solutions to a system of partial differential equations modelling chemotaxis*, in "Trans. Amer. Math. Soc.", 1992, vol. 329, n$^{o}$ 2, p. 819–824, http://dx.doi.org/10.2307/2153966

[128] L. KANTOROVITCH.*On the translocation of masses*, in "C. R. (Doklady) Acad. Sci. URSS (N.S.)", 1942, vol. 37, p. 199–201

[129] E. KLANN.*A Mumford-Shah-Like Method for Limited Data Tomography with an Application to Electron Tomography*, in "SIAM J. Imaging Sciences", 2011, vol. 4, n$^{o}$ 4, p. 1029–1048

[130] J.-M. LASRY, P.-L. LIONS.*Mean field games*, in "Jpn. J. Math.", 2007, vol. 2, n$^{o}$ 1, p. 229–260, http://dx.doi.org/10.1007/s11537-007-0657-8

[131] J. LASSERRE.*Global Optimization with Polynomials and the Problem of Moments*, in "SIAM Journal on Optimization", 2001, vol. 11, n$^{o}$ 3, p. 796-817

[132] J. LELLMANN, D. A. LORENZ, C. SCHÖNLIEB, T. VALKONEN.*Imaging with Kantorovich-Rubinstein Discrepancy*, in "SIAM J. Imaging Sciences", 2014, vol. 7, n$^{o}$ 4, p. 2833–2859

[133] A. S. LEWIS.*Active sets, nonsmoothness, and sensitivity*, in "SIAM Journal on Optimization", 2003, vol. 13, n$^{o}$ 3, p. 702–725

[134] B. LI, F. HABBAL, M. ORTIZ.*Optimal transportation meshfree approximation schemes for Fluid and plastic Flows*, in "Int. J. Numer. Meth. Engng 83:1541–579", 2010, vol. 83, p. 1541–1579

[135] G. LOEPER.*A fully nonlinear version of the incompressible Euler equations: the semigeostrophic system*, in "SIAM J. Math. Anal.", 2006, vol. 38, n$^{o}$ 3, p. 795–823 (electronic)

[136] G. LOEPER, F. RAPETTI.*Numerical solution of the Monge-Ampére equation by a Newton's algorithm*, in "C. R. Math. Acad. Sci. Paris", 2005, vol. 340, n$^{o}$ 4, p. 319–324

[137] D. LOMBARDI, E. MAITRE.*Eulerian models and algorithms for unbalanced optimal transport*, in "Preprint hal-00976501", 2013

[138] C. LÉONARD.*A survey of the Schrödinger problem and some of its connections with optimal transport*, in "Discrete Contin. Dyn. Syst.", 2014, vol. 34, n$^{o}$ 4, p. 1533–1574, http://dx.doi.org/10.3934/dcds.2014.34.1533

[139] J. MAAS, M. RUMPF, C.-B. SCHÖNLIEB, S. SIMON.*A generalized model for optimal transport of images including dissipation and density modulation*, in "Arxiv preprint", 2014

[140] S. G. MALLAT.*A wavelet tour of signal processing*, Third, Elsevier/Academic Press, Amsterdam, 2009

[141] B. MAURY, A. ROUDNEFF-CHUPIN, F. SANTAMBROGIO.*A macroscopic crowd motion model of gradient flow type*, in "Math. Models Methods Appl. Sci.", 2010, vol. 20, n$^{o}$ 10, p. 1787–1821, http://dx.doi.org/10.1142/S0218202510004799

[142] M. I. MILLER, A. TROUVÉ, L. YOUNES.*Geodesic Shooting for Computational Anatomy*, in "Journal of Mathematical Imaging and Vision", March 2006, vol. 24, n$^{o}$ 2, p. 209–228, http://dx.doi.org/10.1007/s10851-005-3624-0

[143] J.-M. MIREBEAU.*Adaptive, Anisotropic and Hierarchical cones of Discrete Convex functions*, in "Preprint", 2014

[144] J.-M. MIREBEAU.*Anisotropic Fast-Marching on Cartesian Grids Using Lattice Basis Reduction*, in "SIAM Journal on Numerical Analysis", 2014, vol. 52, n$^{o}$ 4, p. 1573-1599

[145] Q. MÉRIGOT.*A multiscale approach to optimal transport*, in "Computer Graphics Forum", 2011, vol. 30, n$^{o}$ 5, p. 1583–1592

[146] Q. MÉRIGOT, É. OUDET.*Handling Convexity-Like Constraints in Variational Problems*, in "SIAM J. Numer. Anal.", 2014, vol. 52, n$^{o}$ 5, p. 2466–2487

[147] N. PAPADAKIS, G. PEYRÉ, É. OUDET.*Optimal Transport with Proximal Splitting*, in "SIAM Journal on Imaging Sciences", 2014, vol. 7, n$^{o}$ 1, p. 212–238 [*DOI :* 10.1137/130920058], http://hal.archives-ouvertes.fr/hal-00816211/

[148] B. PASS, N. GHOUSSOUB.*Optimal transport: From moving soil to same-sex marriage*, in "CMS Notes", 2013, vol. 45, p. 14–15

[149] B. PASS.*Uniqueness and Monge Solutions in the Multimarginal Optimal Transportation Problem*, in "SIAM Journal on Mathematical Analysis", 2011, vol. 43, n° 6, p. 2758-2775

[150] B. PERTHAME, F. QUIROS, J. L. VAZQUEZ.*The Hele-Shaw Asymptotics for Mechanical Models of Tumor Growth*, in "Archive for Rational Mechanics and Analysis", 2014, vol. 212, n° 1, p. 93-127, http://dx.doi.org/10.1007/s00205-013-0704-y

[151] J. PETITOT.*The neurogeometry of pinwheels as a sub-riemannian contact structure*, in "Journal of Physiology-Paris", 2003, vol. 97, n° 23, p. 265–309

[152] G. PEYRÉ.*Texture Synthesis with Grouplets*, in "Pattern Analysis and Machine Intelligence, IEEE Transactions on", April 2010, vol. 32, n° 4, p. 733–746

[153] B. PICCOLI, F. ROSSI.*Generalized Wasserstein distance and its application to transport equations with source*, in "Archive for Rational Mechanics and Analysis", 2014, vol. 211, n° 1, p. 335–358

[154] C. POON.*Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames*, in "Applied and Computational Harmonic Analysis", 2015

[155] H. RAGUET, J. FADILI, G. PEYRÉ.*A Generalized Forward-Backward Splitting*, in "SIAM Journal on Imaging Sciences", 2013, vol. 6, n° 3, p. 1199–1226 [*DOI :* 10.1137/120872802], http://hal.archives-ouvertes.fr/hal-00613637/

[156] J.-C. ROCHET, P. CHONÉ.*Ironing, Sweeping and multi-dimensional screening*, in "Econometrica", 1998

[157] L. RUDIN, S. OSHER, E. FATEMI.*Nonlinear total variation based noise removal algorithms*, in "Physica D: Nonlinear Phenomena", 1992, vol. 60, n° 1, p. 259–268, http://dx.doi.org/10.1016/0167-2789(92)90242-F

[158] O. SCHERZER, M. GRASMAIR, H. GROSSAUER, M. HALTMEIER, F. LENZEN.*Variational Methods in Imaging*, Springer, 2008

[159] T. SCHMAH, L. RISSER, F.-X. VIALARD.*Left-Invariant Metrics for Diffeomorphic Image Registration with Spatially-Varying Regularisation*, in "MICCAI (1)", 2013, p. 203-210

[160] T. SCHMAH, L. RISSER, F.-X. VIALARD.*Diffeomorphic image matching with left-invariant metrics*, in "Fields Institute Communications series, special volume in memory of Jerrold E. Marsden", January 2014

[161] J. SOLOMON, F. DE GOES, G. PEYRÉ, M. CUTURI, A. BUTSCHER, A. NGUYEN, T. DU, L. GUIBAS.*Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains*, in "ACM Transaction on Graphics, Proc. SIGGRAPH'15", 2015, to appear

[162] R. TIBSHIRANI.*Regression shrinkage and selection via the Lasso*, in "Journal of the Royal Statistical Society. Series B. Methodological", 1996, vol. 58, n° 1, p. 267–288

[163] A. TROUVÉ, F.-X. VIALARD.*Shape splines and stochastic shape evolutions: A second order point of view*, in "Quarterly of Applied Mathematics", 2012

[164] S. VAITER, M. GOLBABAEE, J. FADILI, G. PEYRÉ.*Model Selection with Piecewise Regular Gauges*, in "Information and Inference", 2015, to appear, http://hal.archives-ouvertes.fr/hal-00842603/

[165] F.-X. VIALARD, L. RISSER, D. RUECKERT, C. COTTER.*Diffeomorphic 3D Image Registration via Geodesic Shooting Using an Efficient Adjoint Calculation*, in "International Journal of Computer Vision", 2012, vol. 97, n$^{o}$ 2, p. 229-241, http://dx.doi.org/10.1007/s11263-011-0481-8

[166] F.-X. VIALARD, L. RISSER.*Spatially-Varying Metric Learning for Diffeomorphic Image Registration: A Variational Framework*, in "Medical Image Computing and Computer-Assisted Intervention MICCAI 2014", Lecture Notes in Computer Science, Springer International Publishing, 2014, vol. 8673, p. 227-234

[167] C. VILLANI.*Topics in optimal transportation*, Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003, vol. 58, xvi+370

[168] C. VILLANI.*Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009, vol. 338, xxii+973, Old and new, http://dx.doi.org/10.1007/978-3-540-71050-9

[169] X.-J. WANG.*On the design of a reflector antenna. II*, in "Calc. Var. Partial Differential Equations", 2004, vol. 20, n$^{o}$ 3, p. 329–341, http://dx.doi.org/10.1007/s00526-003-0239-4

[170] B. WIRTH, L. BAR, M. RUMPF, G. SAPIRO.*A continuum mechanical approach to geodesics in shape space*, in "International Journal of Computer Vision", 2011, vol. 93, n$^{o}$ 3, p. 293–318

[171] J. WRIGHT, Y. MA, J. MAIRAL, G. SAPIRO, T. S. HUANG, S. YAN.*Sparse representation for computer vision and pattern recognition*, in "Proceedings of the IEEE", 2010, vol. 98, n$^{o}$ 6, p. 1031–1044

# Team OURAGAN

## OUtils de Résolution Algébriques pour la Géométrie et ses 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
**Paris**

THEME
**Algorithmics, Computer Algebra and Cryptology**

# Table of contents

# Team OURAGAN

*Creation of the Team: 2012 January 01*

**Keywords:**

**Computer Science and Digital Science:**

A4.3. - Cryptography
A4.3.1. - Public key cryptography
A4.3.2. - Secret key cryptography
A4.3.3. - Cryptographic protocols
A4.3.4. - Quantum Cryptography
A7.1. - Algorithms
A7.1.4. - Quantum algorithms
A8.1. - Discrete mathematics, combinatorics
A8.3. - Geometry, Topology
A8.4. - Computer Algebra
A8.5. - Number theory
A8.10. - Computer arithmetic

**Other Research Topics and Application Domains:**

B5.6. - Robotic systems
B9.5.1. - Computer science
B9.5.2. - Mathematics

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Fabrice Rouillier [Team leader, Inria, Senior Researcher, HDR]
Razvan Barbulescu [CNRS, Researcher]

**Faculty Members**

Elisha Falbel [Sorbonne Université, Professor, HDR]
Antonin Guilloux [Sorbonne Université, Associate Professor, HDR]
Antoine Joux [Sorbonne Université, Professor, HDR]
Pierre-Vincent Koseleff [Sorbonne Université, Associate Professor, HDR]

**PhD Students**

Thomas Espitau [Sorbonne Université]
Mahya Mehrabdollahei [Inria, from Oct 2018]
Natalia Kharchenko [Sorbonne Université]
Sudarshan Shinde [Sorbonne Université]
Robin Timsit [Sorbonne Université]

**Post-Doctoral Fellows**

Anand Kumar Naranayan [Sorbonne Université]
Irene Pasquinelli [Sorbonne Université]

**Administrative Assistant**

Laurence Bourcier [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

OURAGAN proposes to focus on the transfer of computational algebraic methods to some related fields (computational geometry, topology, number theory, etc.) and some carefully chosen application domains (robotics, control theory, evaluation of the security of cryptographic systems, etc.), which implies working equally on the use (modeling, know - how) and on the development of new algorithms. The latest breakthrough developments and applications where algebraic methods are currently decisive remain few and very targeted. We wish to contribute to increase the impact of these methods but also the number of domains where the use of computational algebraic methods represent a significant added value. This transfer-oriented positioning does not imply to stop working on the algorithms, it simply sets the priorities.

An original aspect of the OURAGAN proposal is to blend into an environment of fundamental mathematics, at the Institut de Mathématiques de Jussieu – Paris Rive Gauche (IMJ-PRG CNRS 7586), and to be cross-functional to several teams (Algebraic Analysis, Complex Analysis and Geometry, Number Theory to name only the main ones), which will be our first source of transfer of computational know-how. The success of this coupling allows to maintain a strong theoretical basis and to measure objectively our transfer activity in the direction of mathematicians (in geometry, topology, number theory, etc.) and to consolidate the presence of Inria in scientific areas among the most theoretical.

We propose two general directions with four particular targets:

- Number Theory
  - Algorithmic Number Theory
- Topology in small dimension
  - Character varieties
  - Knot theory
  - Computational geometry

These actions come, of course, in addition to the study and development of a common set of core elements of

- Basic theory and algorithms in algebra and geometry [Led by Antoine Joux and Fabrice Rouillier].

This core activity is the invention and study of fundamental algebraic algorithms and objects that can be grouped into 2 categories: algorithms designed to operate on finite fields and algorithms running on fields of characteristic 0; with 2 types of computational strategies: the exactness and the use of approximate arithmetic (but with exact results). This mix also installs joint studies between the various axes and is an originality of the project-team. For example many kinds of arithmetic tools around algebraic numbers have to face to similar theoretical problems such as finding a good representation for a number field; almost all problems related to the resolution of algebraic systems will reduce to the study of varieties in small dimension and in particular, most of the time, to the effective computation of the topology of curves and surfaces, or the certified drawing of non algebraic function over an algebraic variety.

The tools and objects developed for research on algorithmic number theory as well as in computational geometry apply quite directly on some selected connected challenging subjects:

- Security of cryptographic systems
- Control theory
- Robotics

These applications will serve for the evaluation of the general tools we develop when used in a different context, in particular their capability to tackle state of the art problems.

## 2.2. Scientific ground

### *2.2.1. Basic computable objects and algorithms*

The basic computable objects and algorithms we study, use, optimize or develop are among the most classical ones in computer algebra and are studied by many people around the world: they mainly focus on basic computer arithmetic, linear algebra, lattices and polynomial system solving.

Our approach for tackling these basic problems, whose solution is important for the work of the whole team, is three-fold. First, for some selected problems, we do propose and develop general algorithms (isolation of real roots of univariate polynomials, parametrizations of solutions of zero-dimensional polynomial systems, solutions of parametric equations, etc.). Second, for a selection of well-known problems, we propose different computational strategies (for example the use of approximate arithmetic to speed up LLL algorithm or root isolators, still certifying the final result). Last, we propose specialized variants of known algorithms optimized for a given problem (for example, dedicated solvers for degenerated bivariate polynomials to be used in the computation of the topology of plane curves).

In the context of OURAGAN, it is important to avoid reinventing the wheel and to re-use wherever possible existing objects and algorithms. The main effort being focused on finding good formulations/modelizations for an efficient use. However, on demand, we will propose implementations at many different levels. For example, for our ongoing work on hybrid strategies for LLL, mixing interval arithmetics and basic linear algebra operations, we have replaced our general reliable multiprecision interval artihmetic package (MPFI [0]) by a dedicated one and managed to save an important factor.

In the activity of OURAGAN, many key objects or algorithms around the resolution of algebraic systems are developed within the team, such as the resolution of polynomials in one variable with real coefficients [77], [66], rational parameterizations of solutions of zero-dimensional systems with rational coefficients [76], [34] or discriminant varieties for solving systems depending on parameters [73].

For our studies in number theory and applications to the security of cryptographic systems, our team works on three categories of basic algorithms: discrete logarithm computations [64] (for example to make progress on the computation of class groups in number fields [56]), network reductions by means of LLL variants [45] and obviously various computations in linear algebra, for example dedicated to *almost sparse* matrices [65].

These two directions of development are linked at several levels. For example, working with number fields, in particular finding good representations of number fields, lead to the same computational problems as working with roots of polynomial systems by means of triangular systems (towers of number fields) or rational parameterizations (unique number field). Making any progress in one direction will probably have direct consequences for almost all the problems we want to tackle.

Several strategies are also shared between these directions such as the use of approximate arithmetic to speed up certified computations. Sometimes these can also lead to improvement for a different purpose (for example computations over the rationals, deeply used in geometry can often be parallelized combining computations in finite fields together with fast Chinese remaindering and modular evaluations).

As single highlighted example of this sharing of tools and strategies, the use of approximate arithmetic [75] is common to the work on LLL [45] (use in the evaluation of the security of cryptographic systems), resolutions of real-world algebraic systems [66] (used in our applications in robotics and control theory), computations of signs of trigonometric expressions used in knot theory [12] or to certified evaluations of dilogarithm functions on an algebraic variety for the computation of volumes of representations in our work in topology [52].

### *2.2.2. Algorithmic Number Theory*

The frontiers between computable objects, algorithms (above section), computational number theory and applications to security of cryptographic systems are very porous. This union of research fields is mainly driven by the algorithmic improvement to solve presumably hard problems relevant to cryptography, such as computation of discrete logarithms, resolution of hard subset-sum problems, decoding of random binary codes

[0] https://gforge.inria.fr/projects/mpfi/

and search for close and short vectors in lattices. While factorization and discrete logarithm problems have a long history in cryptography, the recent post-quantum cryptosystems introduce a new variety of presumably hard problems/objects/algorithms with cryptographic relevance: the shortest vector problem (SVP), the closest vector problem (CVP) or the computation of isogenies between elliptic curves, especially in the supersingular case.

Solving the discrete logarithm problem in finite fields is a key question for the security of Diffie-Hellman based crypto and was the focus of a lot of academic research over the past 40 years. It is one of the expertise domain in the OURAGAN team.

Members of OURAGAN started working on the topic of discrete logarithms around 1998, with several computation records that were announced on the NMBRTHRY mailing list. In large characteristic, especially for the case of prime fields, the best current method is the number field sieve (NFS) algorithm. In particular, they published the first NFS based record computation [63]. Despite huge practical improvements, the prime field case algorithm hasn't really changed since that first record. Around the same time, we also presented small characteristic computation record based on simplifications of the Function Field Sieve (FFS) algorithm [62].

In 2006, important changes occurred concerning the FFS and NFS algorithms, indeed, while the algorithms only covered the extreme case of constant characteristic and constant extension degree, two papers extended their ranges of applicability to all finite fields. At the same time, this permitted a big simplification of the FFS, removing the need for function fields.

Starting from 2012, new results appeared in small characteristic. Initially based on a simplification of the 2006 result, they quickly blossomed into the Frobenial representation methods, with quasi-polynomial time complexity [28], [64], [57]. Recent progress were also made in larger characteristic [30], [29], [27], [26].

An interesting side-effect of this research was the need to revisit the key sizes of pairing-based cryptography. This type of cryptography is also a topic of interest for OURAGAN. In particular, it was introduced in 2000 [61]. Recent re-evaluation of the necessary key size [26], making use of the overview of the possible discrete logarithm constructions are discussed [25].

The computations of *class groups in number fields* has strong links with the computations of discrete logarithms or factorizations using the NFS (number field sieve) strategy which as the name suggests is based on the use of number fields. Roughly speaking, the NFS algorithm uses two number fields and the strategy consists in choosing number fields with small sized coefficients in their definition polynomials. On the contrary, in class group computations, there is a single number field, which is clearly a simplification, but this field is given as input by some fixed definition polynomial. Obviously, the degree of this polynomial as well as the size of its coefficients are both influencing the complexity of the computations so that finding other polynomials representing the same class group but with a better characterization (degree or coefficient's sizes) is a mathematical problem with direct practical consequences. We proposed a method to address the problem in [56], but many issues remain open.

Computing generators of principal ideals of cyclotomic fields is also strongly related to the computation of class groups in number fields. Ideals in cyclotomic fields are used in a number of recent public-key cryptosystems. Among the difficult problems that ensure the safety of these systems, there is one that consists in finding a small generator, if it exists, of an ideal. The case of cyclotomic fields is considered in [33].

We also use the computations of class numbers to search for examples and counter-examples for mathematical conjectures. For example a study of cyclic cubic fields [26] allowed to progress in the study of Greenberg's conjecture [0].

Another consecrated problem in algorithmic number theory is smoothness testing, i.e. given an integer, decide if all its prime factors are smaller than a given bound. The only subexponential algorithm for this is H. Lenstra's elliptic curve method. Many of the families of elliptic curves here were found (according to the authors) by

[0] R. Greenberg, « On the Iwasawa invariants of totally real number fields », American J. of Math., vol. 98, 1976, p. 263-284.

ad-hoc methods. We introduced a new point of view which allows to make rapidly a finite list of families which are guaranteed to contain the good families for the elliptic curve method of factorization [31].

### *2.2.3. Topology in small dimension*

#### *2.2.3.1. Character varieties*

There is a tradition of using computations and software to study and understand the topology of small dimensional manifolds, going back at least to Thurston's works (and before him, Riley's pionering work). The underlying philosophy of these tools is to build combinatorial models of manifolds (for example, the torus is often described as a square with an identification of the sides). For dimension 2, 3, 4, this approach is relevant and effective. In the team OURAGAN, we focus on the dimension 3, where the manifolds are modelized by a finite numbers of tetrahedra with identification of the faces. The software SnapPy [0] implements this strategy and is regularly used as a starting point in our work. Along the same philosophy of implementation, we can also cite Regina [0]. A specific trait of SnapPy is that it focuses on hyperbolic structures on the 3-dimensional manifolds. This setting is the object of a huge amount of theoretical work that were used to speed up computations. For example, some Newton methods were implemented without certification for solving a system of equations, but the theoretical knowledge of the uniqueness of the solution made this implementation efficient enough for the target applications. In recent years, in part under the influence of our team [0], more attention has been given to certified computations and now this is implemented in SnapPy.

This philosophy (modelization of manifolds by quite simple combinatoric models to compute such complicated objects as representations of the fundamental group) was applied in a pioneering work of Falbel[5] when he begins to look for another type of geometry on 3-dimensional manifolds (called CR-spherical geometry). From a computational point of view, this change of objectives was a jump in the unknown: the theoretical justification for the computations were missing, and the number of variables of the systems were multiplied by four. So instead of a relatively small system that could be tackled by Newton methods and numerical approximations, we had to deal with/study (were in front of) relatively big systems (the smallest example being 8 variables of degree 6) with no a priori description of the solutions. This input from OURAGAN was needed and proved to be useful.

Still, the computable objects that appear from the theoretical study are very often outside the reach of automated computations and are to be handled case by case. A few experts around the world have been tackling this kind of computations (Dunfield, Goerner, Heusener, Porti, Tillman, Zickert) and the main current achievement is the *Ptolemy module*[0] for SnapPy.

From these early computational needs, topology in small dimension has historically been the source of collaboration with the IMJ-PRG laboratory. At the beginning, the goal was essentially to provide computational tools for finding geometric structures in triangulated 3-dimensional varieties. Triangulated varieties can be topologically encoded by a collection of tetrahedra with gluing constraints (this can be called a triangulation or mesh, but it is not an approximation of the variety by simple structures, rather a combinatorial model). Imposing a geometric structure on this combinatorial object defines a number of constraints that we can translate into an algebraic system that we then have to solve to study geometric structures of the initial variety, for example in relying on solutions to study representations of the fundamental group of the variety. For these studies, a large part of the computable objects or algorithms we develop are required, from the algorithms for univariate polynomials to systems depending on parameters. It should be noted that most of the computational work lies in the modeling of problems [32] (see [4]) that have strictly no chance to be solved by blindly running the most powerful black boxes: we usually deal here with systems that have 24 to 64 variables, depend on 4 to 8 parameters and with degrees exceeding 10 in each variable. With an ANR [0] funding on the subject, the progress that we did [48](see [4]) were (much) more significant than expected. In particular, we have introduced new

[0] https://www.math.uic.edu/t3m/SnapPy/
[0] https://regina-normal.github.io
[0] as part of the CURVE project
[0] https://www.math.uic.edu/t3m/SnapPy/ptolemy.html
[0] ANR project Structures Géométriques et Triangulations

computable objects with an immediate theoretical meaning (let us say rather with a theoretical link established with the usual objects of the domain), namely, the so-called *deformation variety*.

Recent developments around Mahler measure [24] lead to the study of new computable objects at a cross-road between geometry and number theory.

*2.2.3.2. Knot theory*

Knot theory is a wide area of mathematics. We are interested in polynomial representations of long knots, that is to say polynomial embeddings $\mathbf{R} \to \mathbf{R}^3 \subset \mathbf{S}^3$. Every knot admits a polynomial representation and a natural question is to determine explicit parameterizations, minimal degree parameterizations. On the other hand we are interested to determine what is the knot of a given polynomial smooth embedding $\mathbf{R} \to \mathbf{R}^3$. These questions involve real algebraic curves. Two-bridge knots (or rational knots) are particularly studied because they are much easier to study. The first 26 knots (except $8_5$) are two-bridge knots. It is proved that every knot is a Chebyshev knot [67], that is to say can be parameterized by a Chebyshev curve $(T_a(t), T_b(t), T_c(t+\varphi))$ where $T_n(t) = \cos(n \arccos t)$ is the $n$-th Chebyshev polynomial of the first kind. Chebyshev knots are polynomial analogues of Lissajous knots that have been studied by Jones, Hoste, Lamm...

Our activity in Knot theory is a bridge between our work in computational geometry (topology and drawing of real space curves) and our work on topology in small dimensions (varieties defined as a knot complement). It was first established that any knot can be parameterized by Chebyshev polynomials, then we have studied the properties of harmonic nodes [69] which then opened the way to effective computations. We were able to give an exhaustive, minimal and certified list of Chebyshev parameterizations of the first rational knots, using blind computations [70]. On the other hand, we propose the identification of Chebyshev knot diagrams ([12]) by developing new certified algorithms for computing trigonometric expressions [71], which was also the subject of Tran Cuong's PhD thesis at UPMC [78]. These works share many tools with our action in visualization and computational geometry.

We made use of Chebyshev polynomials so as Fibonacci polynomials which are families of orthogonal polynomials. Considering the Alexander-Conway polynomials as continuant polynomials in the Fibonacci basis, we were able to give a partial answer to Hoste's conjecture on the roots of Alexander polynomials of alternating knots [68].

We study the lexicographic degree of the two-bridge knots, that is to say the minimal (multi)degree of a polynomial representation of a $N$-crossing two-bridge knot. We show that this degree is $(3, b, c)$ with $b + c = 3N$. We have determined the lexicographic degree of the first 362 first two-bridge knots with 12 crossings or fewer [39].Minimal degrees are available [0]. These results make use of the braid theoretical approach developped by Y. Orevkov to study real plane curves and the use of real pseudoholomorphic curves ([2]), the slide isotopies on trigonal diagrams, namely those that never increase the number of crossings [38].

*2.2.3.3. Visualization and Computational Geometry*

The drawing of algebraic curves and surfaces is a critical action in OURAGAN since it is a key ingredient in numerous developments. For example, a certified plot of a discriminant variety could be the only admissible answer that can be proposed for engineering problems that need the resolution of parametric algebraic systems: this variety (and the connected components of its counter part) defines a partition of the parameter's space in regions above which the solutions are numerically stable and topologically simple.

For our action in Algorithmic Geometry, we are associated with the GAMBLE EPI (Inria Nancy Grand Est) with the aim of developing computational techniques for the study, plotting and topology of real algebraic curves and surfaces. The work involves the development of effective methods of resolution of algebraic systems with 2 or 3 variables (see [1] for example) which are basic engines for computing the topology [74], [43] / or plotting.

[0] https://webusers.imj-prg.fr/~pierre-vincent.koseleff/knots/2bk-lexdeg.html

# 3. Research Program

## 3.1. Basic computable objects and algorithms

The development of basic computable objects is somehow *on demand* and depends on all the other directions. However, some critical computations are already known to be bottlenecks and are sources of constant efforts.

Computations with algebraic numbers appear in almost all our activities: when working with number fields in our work in algorithmic number theory as well as in all the computations that involve the use of solutions of zero-dimensional systems of polynomial equations. Among the identified problems: finding good representations for single number fields (optimizing the size and degree of the defining polynomials), finding good representations for towers or products of number fields (typically working with a tower or finding a unique good extension), efficiently computing in practice with number fields (using certified approximation vs working with the formal description based on polynomial arithmetics). Strong efforts are currently done in the understanding of the various strategies by means of tight theoretical complexity studies [43], [72], [35] and many other efforts will be required to find the right representation for the right problem in practice. For example, for isolating critical points of plane algebraic curves, it is still unclear (at least the theoretical complexity cannot help) that an intermediate formal parameterization is more efficient than a triangular decomposition of the system and it is still unclear that these intermediate computations could be dominated in time by the certified final approximation of the roots.

## 3.2. Algorithmic Number Theory

Concerning algorithmic number theory, the main problems we will be considering in the coming years are the following:

- *Number fields.* We will continue working on the problems of class groups and generators. In particular, the existence and accessibility of *good* defining polynomials for a fixed number field remain very largely open. The impact of better polynomials on the algorithmic performance is a very important parameter, which makes this problem essential.
- *Lattice reduction.* Despite a great amount of work in the past 35 years on the LLL algorithm and its successors, many open problems remain. We will continue the study of the use of interval arithmetic in this field and the analysis of variants of LLL along the lines of the *Potential*-LLL which provides improved reduction comparable to BKZ with a small block size but has better performance.
- *Elliptic curves and Drinfeld modules.* The study of elliptic curves is a very fruitful area of number theory with many applications in crypto and algorithms. Drinfeld modules are “cousins” of elliptic curves which have been less explored in the algorithm context. However, some recent advances [44] have used them to provide some fast sophisticated factoring algorithms. As a consequence, it is natural to include these objects in our research directions.

## 3.3. Topology in small dimension

### *3.3.1. Character varieties*

The brute force approach to computable objects from topology of small dimension will not allow any significant progress. As explained above, the systems that arise from these problems are simply outside the range of doable computations. We still continue the work in this direction by a four-fold approach, with all three directions deeply inter-related. First, we focus on a couple of especially meaningful (for the applications) cases, in particular the 3-dimensional manifold called Whitehead link complement. At this point, we are able to make steps in the computation and describe part of the solutions [48], [55]; we hope to be able to complete the computation using every piece of information to simplify the system. Second, we continue the theoretical work to understand more properties of these systems [46]. These properties may prove how useful for the mathematical understanding is the resolution of such systems - or at least the extraction of

meaningful information. This approach is for example carried on by Falbel and his work on configuration of flags [49], [51]. Third, we position ourselves as experts in the know-how of this kind of computations and natural interlocutors for colleagues coming up with a question on such a computable object [53], [55]. This also allows us to push forward the kind of computation we actually do and make progress in the direction of the second point. We are credible interlocutors because our team has the blend of theoretical knowledge and computational capabilities that grants effective resolutions of the problems we are presented. And last, we use the knowledge already acquired to pursue our theoretical study of the CR-spherical geometry [42], [50], [47].

Another direction of work is the help to the community in experimental mathematics on new objects. It involves downsizing the system we are looking at (for example by going back to systems coming from hyperbolic geometry and not CR-spherical geometry) and get the most out of what we can compute, by studying new objects. An example of this research direction is the work of Guilloux around the volume function on deformation varieties. This is a real-analytic function defined on the varieties we specialized in computing. Being able to do effective computations with this function led first to a conjecture [52]. Then, theoretical discussions around this conjecture led to a paper on a new approach to the Mahler measure of some 2-variables polynomials [54]. In turn, this last paper gave a formula for the Mahler measure in terms of a function akin to the volume function applied at points in an algebraic variety whose moduli of coordinates are 1. The OURAGAN team has the expertise to compute all the objects appearing in this formula, opening the way to another area of application. This area is deeply linked with number theory as well as topology of small dimension. It requires all the tools at disposition within OURAGAN.

### 3.3.2. Knot theory

We will carry on the exhaustive search for the lexicographic degrees for the rational knots. They correspond to trigonal space curves: computations in the braid group $B_3$, explicit parametrization of trigonal curves corresponding to "dessins d'enfants", etc. The problem seems much more harder when looking for more general knots.

On the other hand, a natural direction would be: given an explicit polynomial space curve, determine the under/over nature of the crossings when projecting, draw it and determine the known knot [0] it is isotopic to.

### 3.3.3. Vizualisation and Computational Geometry

As mentioned above, the drawing of algebraic curves and surfaces is a critical action in OURAGAN since it is a key ingredient in numerous developments. In some cases, one will need a fully certified study of the variety for deciding existence of solutions (for example a region in a robot's parameter's space with solutions to the DKP above or deciding if some variety crosses the unit polydisk for some stability problems in control-theory), in some other cases just a partial but certified approximation of a surface (path planning in robotics, evaluation of non algebraic functions over an algebraic variety for volumes of knot complements in the study of character varieties).

On the one hand, we will contribute to general tools like ISOTOP [0] under the supervision of the GAMBLE project-team and, on the other hand, we will propose ad-hoc solutions by gluing some of our basic tools (problems of high degrees in robust control theory). The priority is to provide a first software that implements methods that fit as most as possible the very last complexity results we got on several (theoretical) algorithms for the computation of the topology of plane curves.

A particular effort will be devoted to the resolution of overconstraint bivariate systems which are useful for the studies of singular points and to polynomials systems in 3 variables in the same spirit : avoid the use of Gröbner basis and propose a new algorithm with a state-of-the-art complexity and with a good practical behavior.

[0] for example the first rational knots are listed at https://team.inria.fr/ouragan/knots
[0] https://isotop.gamble.loria.fr

In parallel, one will have to carefully study the drawing of graphs of non algebraic functions over algebraic complex surfaces for providing several tools which are useful for mathematicians working on topology in small dimension (a well known example is the drawing of amoebia, a way of representing a complex curve on a sheet of paper).

# 4. Application Domains

## 4.1. Security of cryptographic systems

The study of the security of asymmetric cryptographic systems comes as an application of the work carried out in algorithmic number theory and revolves around the development and the use of a small number of general purpose algorithms (lattice reduction, class groups in number fields, discrete logarithms in finite fields, ...). For example, the computation of generators of principal ideals of cyclotomic fields can be seen as one of these applications since these are used in a number of recent public key cryptosystems.

The cryptographic community is currently very actively assessing the threat coming for the development of quantum computers. Indeed, such computers would permit tremendous progresses on many number theoretic problems such as factoring or discrete logarithm computations and would put the security of current cryptosystem under a major risk. For this reason, there is a large global research effort dedicated to finding alternative methods of securing data. In particular, the US standardization agency called NIST has recently launched a standardization process around this issue. In this context, OURAGAN is part of the competition and has submitted a candidate, also published in [13]. This method is based on number-theoretic ideas involving a new presumably difficult problem concerning the Hamming distance of integers modulo large numbers of Mersenne.

## 4.2. Robotics

Algebraic computations have tremendously been used in Robotics, especially in kinematics, since the last quarter of the 20th century. For example, one can cite different proofs for the 40 possible solutions to the direct kinematics problem for steward platforms and companion experiments based on Gröbner basis computations. On the one hand, hard general kinematics problems involve too many variables for pure algebraic methods to be used in place of existing numerical or semi-numerical methods everywhere and everytime, and on the other hand, for some quite large classes, global algebraic studies allow to propose exhaustive classifications that cannot be reached by other methods.

Robotics is a long-standing collaborative work with LS2N (Laboratory of Numerical Sciences of Nantes). Work has recently focused on the offline study of mechanisms, mostly parallel, their singularities or at least some types of singularities (cuspidals robots: cusps in the workspace).

For most parallel or serial manipulators, pose variables and joints variables are linked by algebraic equations and thus lie an algebraic variety. The two-kinematics problems (the direct kinematics problem - DKP- and the inverse kinematics problem - IKP) consist in studying the preimage of the projection of this algebraic variety onto a subset of unknowns. Solving the DKP remains to computing the possible positions for a given set of joint variables values while solving the IKP remains to computing the possible joints variables values for a given position. Algebraic methods have been deeply used in several situations for studying parallel and serial mechanisms, but finally their use stays quite confidential in the design process. Cylindrical Algebraic Decomposition coupled with variable's eliminations by means of Gröbner based computations can be used to model the workspace, the joint space and the computation of singularities. On the one hand, such methods suffer immediately when increasing the number of parameters or when working with imprecise data. On the other hand, when the problem can be handled, they might provide full and exhaustive classifications. The tools we use in that context [41] [40] ( [58], [60], [59]) depend mainly on the resolution of parameter-based systems and therefore of study-dependent curves or flat algebraic surfaces (2 or 3 parameters), thus joining our thematic *Algorithmic Geometry*.

## 4.3. Control theory

Many problems in control theory have been studied using general exact polynomial solvers in the past. One can cite the famous Routh-Hurwitz criterion (late 19th century) for the stability of a linear time invariant (LTI) control system and its relation with Sturm sequences and Cauchy index. However most of the strategies used were involving mostly tools for univariate polynomials and then tried to tackle multivariate problems recursively with respect to the variables. More recent work are using a mix of symbolic/numeric strategies, using semi-definite programming for classes of optimization problems or homotopy methods for some algebraic problems, but still very few practical experiments are currently involving certified algebraic using general solvers for polynomial equations.

Our work in control theory is a recent activity and it is done in collaboration with a group of specialists, the GAIA team, Inria Lille-Nord Europe. We started with a well-known problem, the study of the stability of differential delay systems and multidimensional systems with an important observation: with a correct modelization, some recent algebraic methods, derived from our work in algorithmic geometry and shared with applications in robotics, now allow some previously impossible computations and lead to a better understanding of the problems to be solved [37], [36]. The field is porous to computer algebra since one finds for a long time algebraic criteria of all kinds but the technology seems blocked on a recursive use of one-variable methods, whereas our approach involves the direct processing of problems into a larger number of variables or variants.

The structural stability of $n$-D discrete linear systems (with $n \geq 2$) is a good source of problems of several kinds ranging from solving univariate polynomials to studying algebraic systems depending on parameters. For example, we have shown 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 deciding whether or not a certain system of polynomial equations has real solutions. The use state-of-the-art computer algebra algorithms to check this last condition, and thus the structural stability of multidimensional systems has been validated in several situations from toy examples with parameters to state-of-the-art examples involving the resolution of bivariate systems.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

- In [24], Antonin Guilloux and Julien Marché propose a closed formula for the Mahler measure of a class of bivariate polynomials with rational coefficients (exact polynomials). This class of polynomials contains A-polynomials of knot complements and the authors express the Mahler Measure of a volume function defined on the vanishing set of the polynomial.

  As computing Mahler measures is a well known challenge in number theory and as computing volumes of knots complements is a critical objective for our research on character varieties, this result make an original bridge between our two main research directions.
- A key encapsulation message named Mersenne-756839 has been submitted at the NIST call for standard on Post-Quantum Cryptography. This submission is a complement to the article [13] presented in three invited lectures by Antoine Joux (JFLI (UMI CNRS) / Tokyo university , Nanyang Technological University, LATtice Crypto and Algorithms conference).
- Our agreement with WATERLOO MAPLE INC. has been reviewed for a two years term in 2018. Out next objective is the diffusion of our new solver for univariate polynomials with real coefficients.

# 6. New Software and Platforms

## 6.1. ISOTOP

*Topology and geometry of planar algebraic curves*

KEYWORDS: Topology - Curve plotting - Geometric computing
FUNCTIONAL DESCRIPTION: Isotop is a Maple software for computing the topology of an algebraic plane curve, that is, for computing an arrangement of polylines isotopic to the input curve. This problem is a necessary key step for computing arrangements of algebraic curves and has also applications for curve plotting. This software has been developed since 2007 in collaboration with F. Rouillier from Inria Paris - Rocquencourt.
NEWS OF THE YEAR: In 2018, an engineer from Inria Nancy (Benjamin Dexheimer) finished the implementation of the web server to improve the diffusion of our software.

- Participants: Luis Penaranda, Marc Pouget and Sylvain Lazard
- Contact: Marc Pouget
- Publications: Rational Univariate Representations of Bivariate Systems and Applications - Separating Linear Forms for Bivariate Systems - On The Topology of Planar Algebraic Curves - New bivariate system solver and topology of algebraic curves - Improved algorithm for computing separating linear forms for bivariate systems - Solving bivariate systems using Rational Univariate Representations - On the topology of planar algebraic curves - On the topology of real algebraic plane curves - Bivariate triangular decompositions in the presence of asymptotes - Separating linear forms and Rational Univariate Representations of bivariate systems
- URL: https://isotop.gamble.loria.fr/

## 6.2. RS

FUNCTIONAL DESCRIPTION: Real Roots isolation for algebraic systems with rational coefficients with a finite number of Complex Roots

- Participant: Fabrice Rouillier
- Contact: Fabrice Rouillier
- URL: https://team.inria.fr/ouragan/software/

## 6.3. A NewDsc

*A New Descartes*
KEYWORD: Scientific computing
FUNCTIONAL DESCRIPTION: Computations of the real roots of univariate polynomials with rational coefficients.

- Authors: Fabrice Rouillier, Alexander Kobel and Michael Sagraloff
- Partner: Max Planck Institute for Software Systems
- Contact: Fabrice Rouillier
- URL: https://anewdsc.mpi-inf.mpg.de

## 6.4. SIROPA

KEYWORDS: Robotics - Kinematics
FUNCTIONAL DESCRIPTION: Library of functions for certified computations of the properties of articulated mechanisms, particularly the study of their singularities

- Authors: Damien Chablat, Fabrice Rouillier, Guillaume Moroz and Philippe Wenger
- Partner: LS2N
- Contact: Guillaume Moroz
- URL: http://siropa.gforge.inria.fr/

## 6.5. MPFI

KEYWORD: Arithmetic

FUNCTIONAL DESCRIPTION: MPFI is a C library based on MPFR and GMP for multi precision floating point arithmetic.

- Contact: Fabrice Rouillier
- URL: http://mpfi.gforge.inria.fr

# 7. New Results

## 7.1. On $SL(3,\mathbb{C})$-representations of the Whitehead link group

In [9], we describe a family of representations in $SL(3,\mathbb{C})$ of the fundamental group $\pi$ of the Whitehead link complement. These representations are obtained by considering pairs of regular order three elements in $SL(3,\mathbb{C})$ and can be seen as factorising through a quotient of $\pi$ defined by a certain exceptional Dehn surgery on the Whitehead link. Our main result is that these representations form an algebraic component of the $SL(3,\mathbb{C})$-character variety of $\pi$.

## 7.2. A simplified approach to rigorous degree 2 elimination in discrete logarithm algorithms

In [10], we revisit the ZigZag strategy of Granger, Kleinjung and Zumbrägel. In particular, we provide a new algorithm and proof for the so-called degree 2 elimination step. This allows us to provide a stronger theorem concerning discrete logarithm computations in small characteristic fields $F_{q^{k_0 k}}$ with k close to $q$ and $k_0$ a small integer. As in the aforementioned paper, we rely on the existence of two polynomials $h_0$ and $h_1$ of degree 2 providing a convenient representation of the finite field $F_{q^{k_0 k}}$.

## 7.3. Computing Chebyshev knot diagrams

A Chebyshev curve $\mathcal{C}(a,b,c,\phi)$ has a parametrization of the form $x(t) = T_a(t); \quad y(t) = T_b(t); z(t) = T_c(t+\phi)$, where $a, b, c$ are integers, $T_n(t)$ is the Chebyshev polynomial of degree $n$ and $\phi \in \mathbb{R}$. When $\mathcal{C}(a,b,c,\phi)$ is nonsingular, it defines a polynomial knot. In [12], we determine all possible knot diagrams when $\phi$ varies. Let $a, b, c$ be integers, $a$ is odd, $(a,b) = 1$, we show that one can list all possible knots $\mathcal{C}(a,b,c,\phi)$ in $O(n^2)$ bit operations, with $n = abc$.

## 7.4. Programmable projective measurement with linear optics

In [8] present a scheme for a universal device which can be programmed by quantum states to perform a chosen projective measurement, and its implementation in linear optics. In particular, our scheme takes a single input system (the input register), and M-1 systems all in a state psi (the program registers), whose role is to encode the measurement direction, and approximates the projective measurement with respect to the state psi on the input system. Importantly the scheme is entirely independent of the measurement basis choice psi. This is done optimally in M, if we demand the input state psi always returns the appropriate outcome, and limits to the ideal projective measurement with M. The size of the linear optical circuit we propose scales as M log M , and requires O(M log M) classical side processing. Our scheme can be viewed as an extension of the swap test to the instance where one state is supplied many times.

## 7.5. Updating key size estimations for pairings

Recent progress on NFS imposed a new estimation of the security of pairings. In [6], we study the best attacks against some of the most popular pairings. It allows us to propose new pairing-friendly curves of 128 bits and 192 bits of security.

## 7.6. How to Securely Compute with Noisy Leakage in Quasilinear Complexity

Since their introduction in the late 90's, side-channel attacks have been considered as a major threat against cryptographic implementations. This threat has raised the need for formal leakage models in which the security of implementations can be proved. At Eurocrypt 2013, Prouff and Rivain introduced the noisy leakage model which has been argued to soundly capture the physical reality of power and electromagnetic leakages. In their work, they also provide the first formal security proof for a masking scheme in the noisy leakage model. However their work has two important limitations: (i) the security proof relies on the existence of a leak-free component, (ii) the tolerated amount of information in the leakage (aka leakage rate) is of $O(1/n)$ where n is the number of shares in the underlying masking scheme. The first limitation was nicely tackled by Duc, Dziembowski and Faust one year later (Eurocrypt 2014). Their main contribution was to show a security reduction from the noisy leakage model to the conceptually simpler random- probing model. They were then able to prove the security of the well-known Ishai-Sahai-Wagner scheme (Crypto 2003) in the noisy leakage model. The second limitation was addressed last year in a paper by Andrychowicz, Dziembowski and Faust (Eurocrypt 2016). The proposed construction achieves security in the strong adaptive probing model with a leakage rate of $O(1/logn)$ at the cost of a $O(n^2logn)$ complexity. we argue that their result can be translated into the noisy leakage model with a leakage rate of $O(1)$ by using secret sharing based on algebraic geometric codes. They further argue that the efficiency of their construction can be improved by a linear factor using packed secret sharing but no details are provided.

In [14], we show how to compute in the presence of noisy leakage with a leakage rate up to $\widetilde{O}(1)$ in complexity $\widetilde{O}(n)$. They use a polynomial encoding allowing quasilinear multiplication based on the fast Number Theoretic Transform (NTT). They first show that the scheme is secure in the random-probing model with leakage rate $O(1/logn)$. Using the reduction by Duc et al. this result can be translated in the noisy leakage model with a $O(1/|F|^2logn)$ leakage rate. However, as in the work of Andrychowicz et al. , our construction also requires $|F| = O(n)$. In order to bypass this issue, we refine the granularity of our computation by considering the noisy leakage model on logical instructions that work on constant-size machine words. we provide a generic security reduction from the noisy leakage model at the logical-instruction level to the random-probing model at the arithmetic level. This reduction allows to prove the security of the construction in the noisy leakage model with leakage rate $\widetilde{O}(1)$).

## 7.7. A New Public-Key Cryptosystem via Mersenne Numbers

In [13], we propose a new public-key cryptosystem whose security is based on the computational intractability of the following problem: Given a Mersenne number $p = 2^n - 1$ where n is a prime, a positive integer h , and two n -bit integers T,R , find two n -bit integers F,G each of Hamming weight at most h such that T = F ·R+G modulo p , under the promise that they exist.

## 7.8. Workspace, Joint space and Singularities of a family of Delta-Like Robot

In [11], we describe the workspace, the joint space and the singularities of a family of delta-like parallel robots by using algebraic tools. The different func- tions of SIROPA library are introduced, which is used to induce an estimation about the complexity in representing the singularities in the workspace and the joint space. A Gröbner based elimination is used to compute the singularities of the manipulator and a Cylindrical Algebraic Decomposition algorithm is used to study the workspace and the joint space. From these algebraic objects, they propose some certified three-dimensional plotting describing the shape of works- pace and of the joint space which will help the engineers or researchers to decide the most suited configuration of the manipulator they should use for a given task. Also, the different parameters associated with the complexity of the serial and parallel singularities are tabulated, which further enhance the selection of the different configuration of the manipulator by comparing the complexity of the singularity equations.

## 7.9. Certified Non-conservative Tests for the Structural Stability of Discrete Multidimensional Systems

In [7], we present new computer algebra based methods for testing the structural stability of n-D discrete linear systems (with $n \geq 2$). More precisely, they 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.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

The objective of our Agrement with WATERLOO MAPLE INC. is to promote software developments to which we actively contribute.

On the one hand, WMI provides man power, software licenses, technical support (development, documentation and testing) for an inclusion of our developments in their commercial products. On the other hand, OURAGAN offers perpetual licenses for the use of the concerned source code.

As past results of this agreement one can cite our C-Library *RS* for the computations of the real solutions zero-dimensional systems or also our collaborative development around the Maple package *DV* for solving parametric systems of equations.

For this term, the agreement covers algorithms developed in areas including but not limited to: 1) solving of systems of polynomial equations, 2) validated numerical polynomial root finding, 3) computational geometry, 4) curves and surfaces topology, 5) parametric algebraic systems, 6) cylindrical algebraic decompositions, 7) robotics applications.

In particular, it covers our collaborative work with some of our partners, especially the Gamble Project-Team - Inria Nancy Grand Est.

# 9. Partnerships and Cooperations

## 9.1. European Initiatives

### *9.1.1. FP7 & H2020 Projects*

Program:H2020-EU.1.1. - EXCELLENT SCIENCE - European Research Council (ERC)

Project acronym: Almacrypt

Project title: Algorithmic and Mathematical Cryptology

Duration: 01/2016 - 12/2010

Coordinator: Antoine Joux

Abstract: Cryptology is a foundation of information security in the digital world. Today's internet is protected by a form of cryptography based on complexity theoretic hardness assumptions. Ideally, they should be strong to ensure security and versatile to offer a wide range of functionalities and allow efficient implementations. However, these assumptions are largely untested and internet security could be built on sand. The main ambition of Almacrypt is to remedy this issue by challenging the assumptions through an advanced algorithmic analysis. In particular, this proposal questions the two pillars of public-key encryption: factoring and discrete logarithms. Recently, the PI contributed to show that in some cases, the discrete logarithm problem is considerably weaker

than previously assumed. A main objective is to ponder the security of other cases of the discrete logarithm problem, including elliptic curves, and of factoring. We will study the generalization of the recent techniques and search for new algorithmic options with comparable or better efficiency. We will also study hardness assumptions based on codes and subset-sum, two candidates for post-quantum cryptography. We will consider the applicability of recent algorithmic and mathematical techniques to the resolution of the corresponding putative hard problems, refine the analysis of the algorithms and design new algorithm tools. Cryptology is not limited to the above assumptions: other hard problems have been proposed to aim at post-quantum security and/or to offer extra functionalities. Should the security of these other assumptions become critical, they would be added to Almacrypt's scope. They could also serve to demonstrate other applications of our algorithmic progress. In addition to its scientific goal, Almacrypt also aims at seeding a strengthened research community dedicated to algorithmic and mathematical cryptology.

## 9.2. International Initiatives

### *9.2.1. Inria International Labs*

*9.2.1.1. Informal International Partners*

- CQT Singapour (UMI CNRS Majulab)
- UFPA - Para -Brésil (José Miguel Veloso)
- Institut Joseph Fourier - Université Grenoble Alpes (Martin Deraux, V. Vitse et Pierre Will)
- Max-Planck-Institut für Informatik - Saarbrücken - Germany (Michael Sagraloff)
- Holon Institute of Technology, Israel (Jeremy Kaminsky)

## 9.3. International Research Visitors

### *9.3.1. Visits of International Scientists*

- Jeremy Kaminsky (Holon Institute of Technology, Israel). 3-months visitor in Ouragan and École Polytechnique (MAX) and École des Mines. Chateaubriand Fellow. Subjects: Control Theory, Algebraic Geometry and Computer Vision.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. Member of the Organizing Committees*

- Antonin Guilloux is a Co-organizer of the International conference Dynamics of Groups Actions (Cetraro, may 2019) [0]
- Antoine Joux co-organized the Sprint Summer School *Post-Scryptum*[0]
- Antoine-Joux co-organized *Crypto in the quantum age (STIAS)*[0]

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Chair of Conference Program Committees*

- Antoine Joux was Program Chair of Africacrypt [0]

[0] http://dynamicsgroupactions.imj-prg.fr/fr/68-2/
[0] https://postscryptum.lip6.fr
[0] http://stias.ac.za/events/workshop-on-cryptography-in-the-quantum-age
[0] http://africacrypt2018.aui.ma/commitees.php

### 10.1.3. Journal

*10.1.3.1. Member of the Editorial Boards*

- Elisha Falbel is a member of the editorial board of *São Paulo Journal of Mathematical Sciences - Springer*
- Antoine Joux is a member of the editorial board of *Designs, Codes and Cryptography*
- Fabrice Rouillier is a member of the editorial board of *Journal of Symbolic Computation*

*10.1.3.2. Reviewer - Reviewing Activities*

- Antonin Guilloux is reviewer in several journals, including Duke Math Journal.
- Razvan Barbulescu is reviewer for several cryptology conferences including Eurocrypt and WAIFI.

### 10.1.4. Invited Talks

- Razvan Barbulescu, Cryptography and algorithmic number theory, june 2018, Caen
- Elisha Falbel, Colloquium Heidelberg, June 2018 -Heidelberg -Alemagne
- Elisha Falbel, Representation varieties and geometric structures in low dimensions , July 2018 - Warwick-UK
- Elisha Falbel, Modern Trends in Differential Geometry, July 2018, Sao Paulo- Brazil
- Antonin Guilloux, Computation in Geometric Topology, December 2017 - Warwick - UK.
- Antonin Guilloux, Mahler Measure and values of L-functions, August 2018 - Copenhagen - Denmark.
- Antoine Joux, JFLI (UMI CNRS) / Tokyo university, May 2018, Tokyo https://jfliwp.prod.lamp.cnrs.fr/2018/04/13/jfli-seminar-on-the-mersenne-cryptosystem/
- Antoine Joux, Invited Lecture at the conference *Lattice crypto and algorithms*, May 2018, Bertinoro, Italy
- Antoine Joux, The Mersenne Cryptosystem, Nanyang University, June 2018, Singapore

### 10.1.5. Research Administration

- Fabrice Rouillier is a member of the scientific commitee of the Indo French Centre for Applied Mathematics

## 10.2. Teaching - Supervision - Juries

### 10.2.1. Teaching

Elisha Falbel : courses in Algebra and Analysis, L1 , Sorbonne Université.

Elisha Falbel : Course in Probabiltés, L3, Polytech

Elisha Falbel : Introduction aux surfaces de Riemann, M1, Sorbonne Université.

Antonin Guilloux: Courses in General Mathematics, L1, Sorbonne Université.

Antonin Guilloux: Chair of the Mathematics in L1 at Sorbonne Université; Lead of the renewing of the mathematical courses in L1 at Sorbonne Université for 2019.

Antonin Guilloux: Course in Hyperbolic geometry and character varieties, M2, Sorbonne Université.

Antoine Joux : Course on Techniques in Cryptography and Cryptanalysis, M2, Parisian Master of Research in Computer Science.

Pierre-Vincent Koseleff : Course on Applied Algebra, L3 for undergraduate students (6th semester), Sorbonne Université.

Pierre-Vincent Koseleff : Préparation à l'agrégation de Mathématiques, M2. General Chair and teacher. Sorbonne Université.

Fabrice Rouillier : Course in Algebraic Computations, M1, 24h, Sorbonne Université.

Fabrice Rouillier : Course in "Agrégation Option - C", M2, 31 heures,Sorbonne Université.

Razvan Barbulescu : part of the Course at MPRI Arithmetic algorithms for cryptology 6h

Razvan Barbulescu : 3 projects of cryptology in Python

Razvan Barbulescu : exercice sessions for Algorithmic and complexity 30h

### *10.2.2. Supervision*

PhD in progress : Thomas Espitau, 09/2016, directed by Antoine Joux

PhD in progress : Natalia Kharchenko, 09/2016, directed by Antoine Joux

PhD in progress : Mahya Mehrabdollahei, 09/2018, directed by Antonin Guilloux and Fabrice Rouillier

PhD in progress : Sudarshan Shinde, 09/2016, directed by Razvan Barbulescu and Pierre-Vincent Koseleff

PhD in progress : Robin Timsit, 09/2015, directed by Elisha Falbel

### *10.2.3. Juries*

- Fabrice Rouillier was reviewer of the PhD of Ruben Becker (Universität des Saarlandes)
- Antonin Guilloux, jury of the PhD thesis of Alexandre Bellis - Etude Topologique du Flot Horocyclique Le cas des surfaces Géométriquement Infnies - Supervisor: Françoise Dal'Bo.

## 10.3. Popularization

### *10.3.1. Internal or external Inria responsibilities*

- Razvan Barbulescu is *chargé médiation* at IMJ-PRG
- Razvan Barbulescu is a member of the steering commitee of the association *Animath*[0]
- Fabrice Rouillier is *chargé de mission médiation* at Inria Paris
- Fabrice Rouillier is a member of the editorial board of *Interstices*
- Fabrice Rouillier is the president of the association *Animath*

### *10.3.2. Interventions*

- Razvan Barbulescu co-organizes the *Alkindi*[0] competition on crytography (50000 participants)

# 11. Bibliography

## Major publications by the team in recent years

[1] Y. BOUZIDI, S. LAZARD, G. MOROZ, M. POUGET, F. ROUILLIER, M. SAGRALOFF.*Solving bivariate systems using Rational Univariate Representations*, in "Journal of Complexity", 2016, vol. 37, p. 34–75 [*DOI :* 10.1016/J.JCO.2016.07.002], https://hal.inria.fr/hal-01342211

[2] E. BRUGALLÉ, P.-V. KOSELEFF, D. PECKER.*On the lexicographic degree of two-bridge knots*, in "Journal Of Knot Theory And Its Ramifications (JKTR)", June 2016, vol. 25, n° 7, 14p., 21 figs [*DOI :* 10.1142/S0218216516500449], https://hal.archives-ouvertes.fr/hal-01084472

[0] http://www.animath.fr
[0] http://concours-alkindi.fr/

[3] E. FALBEL, A. GUILLOUX.*Dimension of character varieties for 3-manifolds*, in "Proceedings of the American Mathematical Society", 2016 [*DOI :* 10.1090/PROC/13394], https://hal.archives-ouvertes.fr/hal-01370284

[4] E. FALBEL, A. GUILLOUX, P.-V. KOSELEFF, F. ROUILLIER, M. THISTLETHWAITE.*Character Varieties For SL(3,C): The Figure Eight Knot*, in "Experimental Mathematics", 2016, vol. 25, n$^{o}$ 2, 17 [*DOI :* 10.1080/10586458.2015.1068249], https://hal.inria.fr/hal-01362208

[5] E. FALBEL, J. WANG.*Branched spherical CR structures on the complement of the figure-eight knot*, in "Michigan Mathematical Journal", 2014, vol. 63, p. 635-667, https://hal.archives-ouvertes.fr/hal-01374789

## Publications of the year

### Articles in International Peer-Reviewed Journal

[6] R. BARBULESCU, S. DUQUESNE.*Updating key size estimations for pairings*, in "Journal of Cryptology", 2018, https://hal.archives-ouvertes.fr/hal-01534101

[7] 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

[8] U. CHABAUD, E. DIAMANTI, D. MARKHAM, E. KASHEFI, A. JOUX.*Optimal quantum-programmable projective measurement with linear optics*, in "Physical Review A", December 2018, https://arxiv.org/abs/1805.02546 - 11 pages, 7 figures [*DOI :* 10.1103/PHYSREVA.98.062318], https://hal.sorbonne-universite.fr/hal-01931757

[9] A. GUILLOUX, P. WILL.*On SL(3,C)-representations of the Whitehead link group*, in "Geometriae Dedicata", 2018, https://arxiv.org/abs/1607.01536 [*DOI :* 10.1007/S10711-018-0404-8], https://hal.archives-ouvertes.fr/hal-01370289

[10] F. GÖLOĞLU, A. JOUX.*A simplified approach to rigorous degree 2 elimination in discrete logarithm algorithms*, in "Mathematics of Computation", 2018, 1, https://hal.archives-ouvertes.fr/hal-01960765

[11] R. JHA, D. CHABLAT, L. BARON, F. ROUILLIER, G. MOROZ.*Workspace, Joint space and Singularities of a family of Delta-Like Robot*, in "Mechanism and Machine Theory", September 2018, vol. 127, p. 73-95 [*DOI :* 10.1016/J.MECHMACHTHEORY.2018.05.004], https://hal.archives-ouvertes.fr/hal-01796066

[12] P.-V. KOSELEFF, D. PECKER, F. ROUILLIER, C. TRAN.*Computing Chebyshev knot diagrams*, in "Journal of Symbolic Computation", 2018, vol. 86, 21, https://arxiv.org/abs/1512.07766 [*DOI :* 10.1016/J.JSC.2017.04.001], https://hal.inria.fr/hal-01232181

### Scientific Books (or Scientific Book chapters)

[13] D. AGGARWAL, A. JOUX, A. PRAKASH, M. SANTHA.*A New Public-Key Cryptosystem via Mersenne Numbers*, in "Advances in Cryptology - CRYPTO 2018 - 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19-23, 2018, Proceedings, Part III", Springer, 2018, p. 459-482, https://hal.archives-ouvertes.fr/hal-01960756

[14] D. GOUDARZI, A. JOUX, M. RIVAIN.*How to Securely Compute with Noisy Leakage in Quasilinear Complexity*, in "Advances in Cryptology - ASIACRYPT 2018 - 24th International Conference on the Theory

and Application of Cryptology and Information Security, Brisbane, QLD, Australia, December 2-6, 2018, Proceedings, Part II", Springer, October 2018, p. 547-574, https://hal.archives-ouvertes.fr/hal-01960745

[15] A. GUILLOUX, S. LANFRANCHI, É. VARCIN.*Mussolini et les mots de la race*, in "La pensée de la race en Italie : du romantisme au fascisme", A. ARAMINI, E. BOVO (editors), Cahiers de la MSHE Ledoux. Archives de l'imaginaire social, Presses universitaires de Franche-Comté, May 2018, p. 169-184, https://hal.archives-ouvertes.fr/hal-01768471

### Other Publications

[16] D. M. ALMEIDA, E. E. FALBEL.*Fat sub-Riemannian symmetric spaces: the nilpotent case*, May 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01791316

[17] R. BARBULESCU, S. SHINDE.*A complete classification of ECM-friendly families using modular curves*, June 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01822144

[18] E. BRUGALLÉ, P.-V. KOSELEFF, D. PECKER.*The lexicographic degree of the first two-bridge knots*, September 2018, https://arxiv.org/abs/1501.06393 - 30 p., 58 fig., 6 tables, https://hal.archives-ouvertes.fr/hal-01108678

[19] D. N. DIATTA, S. DIATTA, F. ROUILLIER, M.-F. ROY, M. SAGRALOFF.*Bounds for polynomials on algebraic numbers and application to curve topology*, October 2018, https://arxiv.org/abs/1807.10622 - working paper or preprint, https://hal.inria.fr/hal-01891417

[20] E. FALBEL, A. GUILLOUX, P. WILL.*Hilbert metric, beyond convexity*, 2018, https://arxiv.org/abs/1804.05317 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01768400

[21] E. FALBEL, M. MACULAN, G. SARFATTI.*Configurations of flags in orbits of real forms*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01779459

[22] E. FALBEL, J. VELOSO.*Flag structures on real 3-manifolds*, April 2018, https://arxiv.org/abs/1804.11096 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01778582

[23] A. GUILLOUX, I. KIM.*Deformation space of discrete groups of SU(2,1) in quaternionic hyperbolic plane*, March 2018, https://arxiv.org/abs/1803.05231 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01736953

[24] A. GUILLOUX, J. MARCHÉ.*Volume function and Mahler measure of exact polynomials*, April 2018, https://arxiv.org/abs/1804.01395 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01758986

## References in notes

[25] R. BARBULESCU.*A brief history of pairings*, in "International Workshop on the Arithmetic of Finite Fields WAIFI 2016", Gand, Belgium, Arithmetic of Finite Fields – WAIFI 2016, Springer, July 2016, vol. 10064, https://hal.archives-ouvertes.fr/hal-01363444

[26] R. BARBULESCU, S. DUQUESNE.*Updating key size estimations for pairings*, in "Journal of Cryptology", December 2018, vol. published online, to appear in print, https://hal.archives-ouvertes.fr/hal-01534101

[27] R. BARBULESCU, P. GAUDRY, A. GUILLEVIC, F. MORAIN.*Improving NFS for the discrete logarithm problem in non-prime finite fields*, in "Eurocrypt 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques", Sofia, Bulgaria, M. FISCHLIN, E. OSWALD (editors), Advances in Cryptology – EUROCRYPT 2015, April 2015, vol. 9056, p. 129-155 [*DOI :* 10.1007/978-3-662-46800-5_6], https://hal.inria.fr/hal-01112879

[28] R. BARBULESCU, P. GAUDRY, A. JOUX, E. THOMÉ.*A heuristic quasi-polynomial algorithm for discrete logarithm in finite fields of small characteristic*, in "Eurocrypt 2014", Copenhagen, Denmark, P. Q. NGUYEN, E. OSWALD (editors), Advances in Cryptology - EUROCRYPT 2014, Springer, May 2014, vol. 8441, p. 1-16 [*DOI :* 10.1007/978-3-642-55220-5_1], https://hal.inria.fr/hal-00835446

[29] R. BARBULESCU, P. GAUDRY, T. KLEINJUNG.*The Tower Number Field Sieve*, in "ASIACRYPT 2015", Auckland, New Zealand, T. IWATA, J. H. CHEON (editors), Advances in cryptology-Asiacrypt 2015, Springer, November 2015, vol. 9453, p. 31-58, https://hal.archives-ouvertes.fr/hal-01155635

[30] R. BARBULESCU, C. PIERROT.*The Multiple Number Field Sieve for Medium and High Characteristic Finite Fields*, in "LMS Journal of Computation and Mathematics", 2014, vol. 17, p. 230–246 [*DOI :* 10.1112/S1461157014000369], https://hal.inria.fr/hal-00952610

[31] R. BARBULESCU, S. SHINDE.*A complete classification of ECM-friendly families using modular curves*, June 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01822144

[32] N. BERGERON, E. FALBEL, A. GUILLOUX.*Tetrahedra of flags, volume and homology of SL(3)*, in "Geometry & Topology Monographs", 2014, vol. 18 [*DOI :* 10.2140/GT.2014.18.1911], https://hal.archives-ouvertes.fr/hal-01370258

[33] J.-F. BIASSE, T. ESPITAU, P.-A. FOUQUE, A. GÉLIN, P. KIRCHNER.*Computing generator in cyclotomic integer rings*, in "36th Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT 2017)", Paris, France, Lecture Notes in Computer Science, April 2017, vol. 10210, p. 60-88 [*DOI :* 10.1007/978-3-319-56620-7_3], https://hal.archives-ouvertes.fr/hal-01518438

[34] Y. BOUZIDI, S. LAZARD, G. MOROZ, M. POUGET, F. ROUILLIER.*Improved algorithm for computing separating linear forms for bivariate systems*, in "ISSAC - 39th International Symposium on Symbolic and Algebraic Computation", Kobe, Japan, July 2014, https://hal.inria.fr/hal-00992634

[35] Y. BOUZIDI, S. LAZARD, G. MOROZ, M. POUGET, F. ROUILLIER, M. SAGRALOFF.*Solving bivariate systems using Rational Univariate Representations*, in "Journal of Complexity", 2016, vol. 37, p. 34–75 [*DOI :* 10.1016/J.JCO.2016.07.002], https://hal.inria.fr/hal-01342211

[36] Y. BOUZIDI, A. QUADRAT, F. ROUILLIER.*Certified Non-conservative Tests for the Structural Stability of Multidimensional Systems*, August 2017, 31, To appear in Multidimensional Systems and Signal Processing, https://link.springer.com/article/10.1007/s11045-018-0596-y, https://hal.inria.fr/hal-01571230

[37] Y. BOUZIDI, F. ROUILLIER.*Certified Algorithms for proving the structural stability of two dimensional systems possibly with parameters*, in "MNTS 2016 - 22nd International Symposium on Mathematical Theory of Networks and Systems", Minneapolis, United States, Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems, July 2016, https://hal.inria.fr/hal-01366202

[38] E. BRUGALLÉ, P.-V. KOSELEFF, D. PECKER.*Untangling trigonal diagrams*, in "Journal Of Knot Theory And Its Ramifications (JKTR)", June 2016, vol. 25, n$^{o}$ 7, 10p., 24 figs [*DOI :* 10.1142/S0218216516500437], https://hal.archives-ouvertes.fr/hal-01084463

[39] E. BRUGALLÉ, P.-V. KOSELEFF, D. PECKER.*The lexicographic degree of the first two-bridge knots*, September 2018, 30 p., 58 fig., 6 tables, submitted, https://hal.archives-ouvertes.fr/hal-01108678

[40] D. CHABLAT, R. JHA, F. ROUILLIER, G. MOROZ.*Non-singular assembly mode changing trajectories in the workspace for the 3-RPS parallel robot*, in "14th International Symposium on Advances in Robot Kinematics", Ljubljana, Slovenia, June 2014, p. 149 - 159, https://hal.archives-ouvertes.fr/hal-00956325

[41] D. CHABLAT, R. JHA, F. ROUILLIER, G. MOROZ.*Workspace and joint space analysis of the 3-RPS parallel robot*, in "ASME 2013 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference", Buffalo, United States, August 2014, vol. Volume 5A, p. 1-10, https://hal.archives-ouvertes.fr/hal-01006614

[42] M. DERAUX, E. FALBEL.*Complex hyperbolic geometry of the figure eight knot*, in "Geometry and Topology", February 2015, vol. 19, p. 237–293 [*DOI :* 10.2140/GT.2015.19.237], https://hal.archives-ouvertes.fr/hal-00805427

[43] D. N. DIATTA, S. DIATTA, F. ROUILLIER, M.-F. ROY, M. SAGRALOFF.*Bounds for polynomials on algebraic numbers and application to curve topology*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01891417

[44] J. DOLISKANI, A. K. NARAYANAN, É. SCHOST.*Drinfeld Modules with Complex Multiplication, Hasse Invariants and Factoring Polynomials over Finite Fields*, in "CoRR", 2017, vol. abs/1712.00669, http://arxiv.org/abs/1712.00669

[45] T. ESPITAU, A. JOUX.*Adaptive precision LLL and Potential-LLL reductions with Interval arithmetic*, in "IACR Cryptology ePrint Archive", 2016, vol. 2016, 528, http://eprint.iacr.org/2016/528

[46] E. FALBEL, A. GUILLOUX.*Dimension of character varieties for 3-manifolds*, in "Proceedings of the American Mathematical Society", 2016 [*DOI :* 10.1090/PROC/13394], https://hal.archives-ouvertes.fr/hal-01370284

[47] E. FALBEL, A. GUILLOUX, P. WILL.*Hilbert metric, beyond convexity*, 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01768400

[48] E. FALBEL, P.-V. KOSELEFF, F. ROUILLIER.*Representations of fundamental groups of 3-manifolds into PGL(3,C): Exact computations in low complexity*, in "Geometriae Dedicata", August 2015, vol. 177, n$^{o}$ 1, 52 [*DOI :* 10.1007/S10711-014-9987-X], https://hal.inria.fr/hal-00908843

[49] E. FALBEL, M. MACULAN, G. SARFATTI.*Configurations of flags in orbits of real forms*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01779459

[50] E. FALBEL, R. SANTOS THEBALDI.*A Flag structure on a cusped hyperbolic 3-manifold with unipotent holonomy*, in "Pacific Journal of Mathematics", 2015, vol. 278, n$^{o}$ 1, p. 51-78, https://hal.archives-ouvertes.fr/hal-00958255

[51] E. FALBEL, J. VELOSO.*Flag structures on real 3-manifolds*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01778582

[52] A. GUILLOUX.*Volume of representations and birationality of peripheral holonomy*, in "Experimental Mathematics", May 2017, https://hal.archives-ouvertes.fr/hal-01370287

[53] A. GUILLOUX, I. KIM.*Deformation space of discrete groups of SU(2,1) in quaternionic hyperbolic plane*, March 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01736953

[54] A. GUILLOUX, J. MARCHÉ.*Volume function and Mahler measure of exact polynomials*, April 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01758986

[55] A. GUILLOUX, P. WILL.*On SL(3,C)-representations of the Whitehead link group*, 2018, To appear in Geom. Ded., https://hal.archives-ouvertes.fr/hal-01370289

[56] A. GÉLIN, A. JOUX.*Reducing number field defining polynomials: an application to class group computations*, in "Algorithmic Number Theory Symposium XII", Kaiserslautern, Germany, LMS Journal of Computation and Mathematics, August 2016, vol. 19, n$^{o}$ A, p. 315–331 [*DOI :* 10.1112/S1461157016000255], https://hal.archives-ouvertes.fr/hal-01362144

[57] F. GÖLOĞLU, A. JOUX.*A Simplified Approach to Rigorous Degree 2 Elimination in Discrete Logarithm Algorithms*, in "IACR Cryptology ePrint Archive", 2018, vol. 2018, 430, https://eprint.iacr.org/2018/430

[58] R. JHA, D. CHABLAT, L. BARON, F. ROUILLIER, G. MOROZ.*Workspace, Joint space and Singularities of a family of Delta-Like Robot*, in "Mechanism and Machine Theory", September 2018, vol. 127, p. 73-95 [*DOI :* 10.1016/J.MECHMACHTHEORY.2018.05.004], https://hal.archives-ouvertes.fr/hal-01796066

[59] R. JHA, D. CHABLAT, F. ROUILLIER, G. MOROZ.*An algebraic method to check the singularity-free paths for parallel robots*, in "International Design Engineering Technical Conferences & Computers and Information in Engineering Conference", Boston, United States, ASME, August 2015, https://hal.archives-ouvertes.fr/hal-01142989

[60] R. JHA, D. CHABLAT, F. ROUILLIER, G. MOROZ.*Workspace and Singularity analysis of a Delta like family robot*, in "4th IFTOMM International Symposium on Robotics and Mechatronics", Poitiers, France, June 2015, https://hal.archives-ouvertes.fr/hal-01142465

[61] A. JOUX.*A one round protocol for tripartite Diffie-Hellman*, in "J. Cryptology", 2004, vol. 17, n$^{o}$ 4, p. 263–276

[62] A. JOUX, R. LERCIER.*The function field sieve is quite special*, in "Algorithmic Number Theory-ANTS V", Lecture Notes in Computer Science, Springer, 2002, vol. 2369, p. 431-445

[63] A. JOUX, R. LERCIER.*Improvements to the general number field sieve for discrete logarithms in prime fields. A comparison with the gaussian integer method*, in "Math. Comput.", 2003, vol. 72, n$^{o}$ 242, p. 953-967

[64] A. JOUX, C. PIERROT.*Improving the Polynomial time Precomputation of Frobenius Representation Discrete Logarithm Algorithms - Simplified Setting for Small Characteristic Finite Fields*, in "20th International Conference on the Theory and Application of Cryptology and Information Security", Kaoshiung, Taiwan,

Lecture Notes in Computer Science, Springer Berlin Heidelberg, December 2014, vol. 8873, p. 378-397 [*DOI :* 10.1007/978-3-662-45611-8_20], https://hal.archives-ouvertes.fr/hal-01213649

[65] A. JOUX, C. PIERROT.*Nearly Sparse Linear Algebra and application to Discrete Logarithms Computations*, in "Contemporary Developments in Finite Fields and Applications ", 2016 [*DOI :* 10.1142/9789814719261_0008], https://hal.inria.fr/hal-01154879

[66] A. KOBEL, F. ROUILLIER, M. SAGRALOFF.*Computing Real Roots of Real Polynomials ... and now For Real!*, in " ISSAC '16 Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation", Waterloo, Canada, ISSAC '16 Proceedings of the ACM on International Symposium on Symbolic and Algebraic Computation, July 2016, 7 [*DOI :* 10.1145/2930889.2930937], https://hal.inria.fr/hal-01363955

[67] P.-V. KOSELEFF, D. PECKER.*Chebyshev Knots*, in "Journal of Knot Theory and Its Ramifications", April 2011, vol. 20, n$^{o}$ 4, p. 575-593 [*DOI :* 10.1142/S0218216511009364], https://hal.archives-ouvertes.fr/hal-00344501

[68] P.-V. KOSELEFF, D. PECKER.*On Alexander–Conway polynomials of two-bridge links*, in "Journal of Symbolic Computation", May 2015, vol. Volume 68, n$^{o}$ 2, p. 215-229, 15p [*DOI :* 10.1016/J.JSC.2014.09.011], https://hal.archives-ouvertes.fr/hal-00538729

[69] P.-V. KOSELEFF, D. PECKER.*Harmonic Knots*, in "Journal Of Knot Theory And Its Ramifications (JKTR)", 2016, vol. 25, n$^{o}$ 13, 18, 18 p., 30 fig. [*DOI :* 10.1142/S0218216516500747], https://hal.archives-ouvertes.fr/hal-00680746

[70] P.-V. KOSELEFF, D. PECKER, F. ROUILLIER.*The first rational Chebyshev knots*, in "Journal of Symbolic Computation", December 2010, vol. 45, n$^{o}$ 12, p. 1341-1358 [*DOI :* 10.1016/J.JSC.2010.06.014], https://hal.archives-ouvertes.fr/hal-00429510

[71] P.-V. KOSELEFF, F. ROUILLIER, C. TRAN.*On the sign of a trigonometric expression*, in "ISSAC ' 15", Bath, United Kingdom, Proceedings of the 2015 ACM on International Symposium on Symbolic and Algebraic Computation, July 2015 [*DOI :* 10.1145/2755996.2756664], https://hal.inria.fr/hal-01200820

[72] S. LAZARD, M. POUGET, F. ROUILLIER.*Bivariate triangular decompositions in the presence of asymptotes*, in "Journal of Symbolic Computation", 2017, vol. 82, p. 123 - 133 [*DOI :* 10.1016/J.JSC.2017.01.004], https://hal.inria.fr/hal-01468796

[73] D. LAZARD, F. ROUILLIER.*Solving Parametric Polynomial Systems*, in "Journal of Symbolic Computation", June 2007, vol. 42, p. 636-667

[74] D. NIANG DIATTA, F. ROUILLIER, M.-F. ROY.*On the computation of the topology of plane curves*, in "International Symposium on Symbolic and Algebraic Computation", Kobe, Japan, K. NABESHIMA (editor), ACM Press, July 2014, p. 130-137 [*DOI :* 10.1145/2608628.2608670], https://hal.archives-ouvertes.fr/hal-00935728

[75] N. REVOL, F. ROUILLIER.*Motivations for an arbitrary precision interval arithmetic and the MPFI library*, in "Reliable Computing", August 2005, vol. 11, n$^{o}$ 4, p. 275-290 [*DOI :* 10.1007/S11155-005-6891-Y], https://hal.inria.fr/inria-00544998

[76] F. ROUILLIER.*Solving zero-dimensional systems through the rational univariate representation*, in "Journal of Applicable Algebra in Engineering, Communication and Computing", 1999, vol. 9, n$^{o}$ 5, p. 433–461

[77] F. ROUILLIER, P. ZIMMERMANN.*Efficient Isolation of Polynomial Real Roots*, in "Journal of Computational and Applied Mathematics", 2003, vol. 162, n$^{o}$ 1, p. 33–50

[78] C. TRAN.*Symbolic computing with the basis of Chebyshev's monic polynomials*, Université Pierre et Marie Curie - Paris VI, October 2015, https://tel.archives-ouvertes.fr/tel-01273287

# Project-Team PARKAS

## Parallélisme de Kahn Synchrone

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Embedded and Real-time Systems**

# Table of contents

# Project-Team PARKAS

*Creation of the Team: 2011 April 01, updated into Project-Team: 2012 January 01*

*The PARKAS team is located at the École normale supérieure*

**Keywords:**

**Computer Science and Digital Science:**

A1.1.1. - Multicore, Manycore
A1.1.3. - Memory models
A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.6. - Concurrent programming
A2.1.9. - Synchronous languages
A2.2.2. - Memory models
A2.2.4. - Parallel architectures
A2.2.5. - Run-time systems
A2.2.6. - GPGPU, FPGA...
A2.2.7. - Adaptive compilation
A2.3. - Embedded and cyber-physical systems
A2.3.1. - Embedded systems
A2.3.2. - Cyber-physical systems
A2.3.3. - Real-time systems
A2.4.3. - Proofs
A3.1.3. - Distributed data
A3.1.8. - Big data (production, storage, transfer)
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.7. - High performance computing

**Other Research Topics and Application Domains:**

B5.2.1. - Road vehicles
B5.2.2. - Railway
B5.2.3. - Aviation
B6.4. - Internet of things
B6.6. - Embedded systems
B9.2.1. - Music, sound
B9.5.1. - Computer science
B9.5.2. - Mathematics

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Timothy Bourke [Inria, Researcher]
Albert Cohen [Inria, Senior Researcher, HDR]
Francesco Zappa Nardelli [Inria, Senior Researcher, HDR]

**Faculty Member**
Marc Pouzet [Team leader, Univ Pierre et Marie Curie, Professor]

**External Collaborator**
Paul Feautrier [Univ de Lyon]

**Technical Staff**
Theophile Bastian [Inria, from Oct 2018 until Nov 2018]
Andi Drebes [Inria, from Oct 2018]
Michael Kruse [Inria, until Jan 2018]
Adilla Susungi [Inria, from Nov 2018]
Nicolas Tollenaere [Inria]
Oleksandr Zinenko [Inria, until Sep 2018]

**PhD Students**
Ulysse Beaugnon [Ecole Normale Supérieure Paris]
Lelio Brun [Ecole Normale Supérieure Paris]
Chandan Reddy Gopal [Inria]
Jie Zhao [Ecole d'ingénieurs]

**Post-Doctoral Fellow**
Guillaume Iooss [Ecole Normale Supérieure Paris]

**Administrative Assistants**
Christine Anocq [Inria]
Nelly Maloisel [Inria, from Dec 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

Research in PARKAS focuses on the design, semantics, and compilation of programming languages which allow going from parallel deterministic specifications to target embedded code executing on sequential or multi-core architectures. We are driven by the ideal of a mathematical and executable language used both to program and simulate a wide variety of systems, including real-time embedded controllers in interaction with a physical environment (e.g., fly-by-wire, engine control), computationally intensive applications (e.g., video), and compilers that produce provably correct and efficient code.

The team bases its research on the foundational work of Gilles Kahn on the semantics of deterministic parallelism, the theory and practice of synchronous languages and typed functional languages, synchronous circuits, modern (polyhedral) compilation, and formal models to prove the correctness of low level code running on weak-memory processors.

To realize our research program, we develop languages (LUCID SYNCHRONE, REACTIVEML, LUCY-N, ZELUS), compilers (PPCG), contributions to open-source projects (isl, LLVM, gcc), tools to study language semantics (Ott) and to test optimization compilers in the presence of threads (cmmtest), and formalizations in Interactive Theorem Provers of language semantics (Vélus, *n*-synchrony, quasi-synchrony). These software projects constitute essential "laboratories": they ground our scientific contributions, guide and validate our research through experimentation, and are an important vehicle for mutually beneficial and long standing collaborations with industry.

# 3. Research Program

## 3.1. Programming Languages for Cyber-Physical Systems

We study the definition of languages for reactive and Cyber-Physical Systems in which distributed control software interacts closely with physical devices. We focus on languages that mix discrete-time and continuous-time; in particular, the combination of synchronous programming constructs with differential equations, relaxed models of synchrony for distributed systems communicating via periodic sampling or through buffers, and the embedding of synchronous features in a general purpose ML language.

The synchronous language SCADE, [0] based on synchronous languages principles, is ideal for programming embedded software and is used routinely in the most critical applications. But embedded design also involves modeling the control software together with its environment made of physical devices that are traditionally defined by differential equations that evolve on a continuous-time basis and approximated with a numerical solver. Furthermore, compilation usually produces single-loop code, but implementations increasingly involve multiple and multi-core processors communicating via buffers and shared-memory.

The major player in embedded design for cyber-physical systems is undoubtedly SIMULINK, [0] with MODELICA[0] a new player. Models created in these tools are used not only for simulation, but also for test-case generation, formal verification, and translation to embedded code. That said, many foundational and practical aspects are not well-treated by existing theory (for instance, hybrid automata), and current tools. In particular, features that mix discrete and continuous time often suffer from inadequacies and bugs. This results in a broken development chain: for the most critical applications, the model of the controller must be reprogrammed into either sequential or synchronous code, and properties verified on the source model have to be reverified on the target code. There is also the question of how much confidence can be placed in the code used for simulation.

We attack these issues through the development of the ZELUS research prototype, industrial collaborations with the SCADE team at ANSYS/Esterel-Technologies, and collaboration with Modelica developers at Dassault-Systèmes and the Modelica association. Our approach is to develop a *conservative extension* of a synchronous language capable of expressing in a single source text a model of the control software and its physical environment, to simulate the whole using off-the-shelf numerical solvers, and to generate target embedded code. Our goal is to increase faithfulness and confidence in both what is actually executed on platforms and what is simulated. The goal of building a language on a strong mathematical basis for hybrid systems is shared with the Ptolemy project at UC Berkeley; our approach is distinguished by building our language on a synchronous semantics, reusing and extending classical synchronous compilation techniques.

Adding continuous time to a synchronous language gives a richer programming model where reactive controllers can be specified in idealized physical time. An example is the so called quasi-periodic architecture studied by Caspi, where independent processors execute periodically and communicate by sampling. We have applied ZELUS to model a class of quasi-periodic protocols and to analyze an abstraction proposed for model-checking such systems.

Communication-by-sampling is suitable for control applications where value timeliness is paramount and lost or duplicate values tolerable, but other applications—for instance, those involving video streams—seek a different trade-off through the use of bounded buffers between processes. We developed the *n*-synchronous model and the programming language LUCY-N to treat this issue.

## 3.2. Efficient Compilation for Parallel and Distributed Computing

We develop compilation techniques for sequential and multi-core processors, and efficient parallel run-time systems for computationally intensive real-time applications (e.g., video and streaming). We study the

[0] http://www.esterel-technologies.com/products/scade-suite
[0] http://www.mathworks.com/products/simulink
[0] https://www.modelica.org

generation of parallel code from synchronous programs, compilation techniques based on the polyhedral model, and the exploitation of synchronous Single Static Assignment (SSA) representations in general purpose compilers.

We consider distribution and parallelism as two distinct concepts.

- Distribution refers to the construction of multiple programs which are dedicated to run on specific computing devices. When an application is designed for, or adapted to, an embedded multiprocessor, the distribution task grants fine grained—design- or compilation-time—control over the mapping and interaction between the multiple programs.
- Parallelism is about generating code capable of efficiently exploiting multiprocessors. Typically this amounts to maing (in)dependence properties, data transfers, atomicity and isolation explicit. Compiling parallelism translates these properties into low-level synchronization and communication primitives and/or onto a runtime system.

We also see a strong relation between the foundations of synchronous languages and the design of compiler intermediate representations for concurrent programs. These representations are essential to the construction of compilers enabling the optimization of parallel programs and the management of massively parallel resources. Polyhedral compilation is one of the most popular research avenues in this area. Indirectly, the design of intermediate representations also triggers exciting research on dedicated runtime systems supporting parallel constructs. We are particularly interested in the implementation of non-blocking dynamic schedulers interacting with decoupled, deterministic communication channels to hide communication latency and optimize local memory usage.

While distribution and parallelism issues arise in all areas of computing, our programming language perspective pushes us to consider four scenarios:

1. designing an embedded system, both hardware and software, and codesign;
2. programming existing embedded hardware with functional and behavioral constraints;
3. programming and compiling for a general-purpose or high-performance, best-effort system;
4. programming large scale distributed, I/O-dominated and data-centric systems.

We work on a multitude of research experiments, algorithms and prototypes related to one or more of these scenarios. Our main efforts focused on extending the code generation algorithms for synchronous languages and on the development of more scalable and widely applicable polyhedral compilation methods.

## 3.3. Validation and Proof of Compilers

Compilers are complex software and not immune from bugs. We work on validation and proof tools for compilers to relate the semantics of executed code and source programs. We develop techniques to formally prove the correctness of compilation passes for synchronous languages (Lustre), and to validate compilation optimization for C code in the presence of threads.

### *3.3.1. Lustre:*

The formal validation of a compiler for a synchronous language (or more generally for a language based on synchronous block diagrams) promises to reduce the likelihood of compiler-introduced bugs, the cost of testing, and also to ensure that properties verified on the source model hold of the target code. Such a validation would be complementary to existing industrial qualifications which certify the development process and not the functional correctness of a compiler. The scientific interest is in developing models and techniques that both facilitate the verification and allow for convenient reasoning over the semantics of a language and the behavior of programs written in it.

### *3.3.2. C/C++:*

The recently approved C11 and C++11 standards define a concurrency model for the C and C++ languages, which were originally designed without concurrency support. Their intent is to permit most compiler and hardware optimizations, while providing escape mechanisms for writing portable, high-performance, low-level code. Mainstream compilers are being modified to support the new standards. A subtle class of compiler bugs is the so-called concurrency compiler bugs, where compilers generate correct sequential code but break the concurrency memory model of the programming language. Such bugs are observable only when the miscompiled functions interact with concurrent contexts, making them particularly hard to detect. All previous techniques to test compiler correctness miss concurrency compiler bugs.

### *3.3.3. Static Analysis of x10*

x10 is an explicit parallel programming language, originally developped by IBM Research. Parallelism is expressed by the `async / finish` construct (a disymetric variant of `fork / join`), and synchronization uses *clocks*, a sophisticated version of barriers. Programs in this language can be analysed at compile time provided their control statements obey the restrictions of of the polyhedral model. The analysis focuses on the extraction of the *happens before* relation of the subject program, and can be used for the detection of races and deadlocks. A first version of this analysis, which did not take clocks into account, was published in 2013. Its extension to clocked programs is a complex problem, which requires the use of a proof assistant, Coq. Work in collaboration with Alain Ketterlin and Eric Violard (Inria Camus) and Tomofumi Yuki (Inria Cairn).

### *3.3.4. Toward a Polynomial Model*

The polyhedral model is a powerful tool for program analysis and verification, autoparallelization, and optimization. However, it can only be applied to a very restricted class of programs : counted loops, affine conditionals and arrays with affine subscripts. The key mathematical result at the bottom of this model is Farkas lemma, which characterizes all affine function non negative on a polyhedron. Recent mathematical results on the *Positiv Stellen Satz* enable a similar characterization for polynomials positive on a semi-algebraic set. Polynomials may be native to the subject code, but also appears as soon as counting is necessary, for instance when a multidimensional array is linearized or when messages are transmitted through a one dimensional channel. Applying the above theorems allows the detection of polynomial dependences and the construction of polynomial schedules, hence the detection of deadlocks. Code generation from a polynomial schedule is the subject of present work. These methods are applied to the language openStream. Work in collaboration with Albert Cohen and Alain Darte (Xilinx).

# 4. Highlights of the Year

## 4.1. Highlights of the Year

Guillaume Baudart was awarded the **ACM SIGBED Paul Caspi Memorial Dissertation Award** for his thesis “A Synchronous Approach to Quasi-Periodic Systems” [27] prepared in the PARKAS Team under the supervision of Marc Pouzet and Timothy Bourke and defended in 2017.

# 5. New Software and Platforms

## 5.1. Cmmtest

FUNCTIONAL DESCRIPTION: Cmmtest is a tool for hunting concurrency compiler bugs. The Cmmtest tool performs random testing of C and C++ compilers against the C11/C++11 memory model. A test case is any well-defined, sequential C program, for each test case, cmmtest:

compiles the program using the compiler and compiler optimisations that are being tested,

runs the compiled program in an instrumented execution environment that logs all memory accesses to global variables and synchronisations,

compares the recorded trace with a reference trace for the same program, checking if the recorded trace can be obtained from the reference trace by valid eliminations, reorderings and introductions.

Cmmtest identified several mistaken write introductions and other unexpected behaviours in the latest release of the gcc compiler. These have been promptly fixed by the gcc developers.

- Participants: Anirudh Kumar, Francesco Zappa Nardelli, Pankaj More, Pankaj Pawan, Pankaj Prateek Kewalramani and Robin Morisset
- Contact: Francesco Zappa Nardelli
- URL: http://www.di.ens.fr/~zappa/projects/cmmtest/

## 5.2. GCC

KEYWORDS: Compilation - Polyhedral compilation
FUNCTIONAL DESCRIPTION: The GNU Compiler Collection includes front ends for C, C++, Objective-C, Fortran, Java, Ada, and Go, as well as libraries for these languages (libstdc++, libgcj,...). GCC was originally written as the compiler for the GNU operating system. The GNU system was developed to be 100

- Participants: Albert Cohen, Feng Li, Nhat Minh Le, Riyadh Baghdadi and Tobias Grosser
- Contact: Albert Cohen
- URL: http://gcc.gnu.org/

## 5.3. Heptagon

KEYWORDS: Compilers - Synchronous Language - Controller synthesis
FUNCTIONAL DESCRIPTION: Heptagon is an experimental language for the implementation of embedded real-time reactive systems. It is developed inside the Synchronics large-scale initiative, in collaboration with Inria Rhones-Alpes. It is essentially a subset of Lucid Synchrone, without type inference, type polymorphism and higher-order. It is thus a Lustre-like language extended with hierchical automata in a form very close to SCADE 6. The intention for making this new language and compiler is to develop new aggressive optimization techniques for sequential C code and compilation methods for generating parallel code for different platforms. This explains much of the simplifications we have made in order to ease the development of compilation techniques.

The current version of the compiler includes the following features: - Inclusion of discrete controller synthesis within the compilation: the language is equipped with a behavioral contract mechanisms, where assumptions can be described, as well as an "enforce" property part. The semantics of this latter is that the property should be enforced by controlling the behaviour of the node equipped with the contract. This property will be enforced by an automatically built controller, which will act on free controllable variables given by the programmer. This extension has been named BZR in previous works. - Expression and compilation of array values with modular memory optimization. The language allows the expression and operations on arrays (access, modification, iterators). With the use of location annotations, the programmer can avoid unnecessary array copies.

- Participants: Adrien Guatto, Brice Gelineau, Cédric Pasteur, Eric Rutten, Gwenaël Delaval, Léonard Gérard and Marc Pouzet
- Partners: UGA - ENS Paris - Inria - LIG
- Contact: Gwenaël Delaval
- URL: http://heptagon.gforge.inria.fr

## 5.4. isl

FUNCTIONAL DESCRIPTION: isl is a library for manipulating sets and relations of integer points bounded by linear constraints. Supported operations on sets include intersection, union, set difference, emptiness check, convex hull, (integer) affine hull, integer projection, transitive closure (and over-approximation), computing the lexicographic minimum using parametric integer programming. It includes an ILP solver based on generalized basis reduction, and a new polyhedral code generator. isl also supports affine transformations for polyhedral compilation, and increasingly abstract representations to model source and intermediate code in a polyhedral framework.

- Participants: Albert Cohen, Sven Verdoolaege and Tobias Grosser
- Contact: Sven Verdoolaege
- URL: http://freshmeat.net/projects/isl

## 5.5. Lem

*lightweight executable mathematics*
FUNCTIONAL DESCRIPTION: Lem is a lightweight tool for writing, managing, and publishing large scale semantic definitions. It is also intended as an intermediate language for generating definitions from domain-specific tools, and for porting definitions between interactive theorem proving systems (such as Coq, HOL4, and Isabelle). As such it is a complementary tool to Ott. Lem resembles a pure subset of Objective Caml, supporting typical functional programming constructs, including top-level parametric polymorphism, datatypes, records, higher-order functions, and pattern matching. It also supports common logical mechanisms including list and set comprehensions, universal and existential quantifiers, and inductively defined relations. From this, Lem generates OCaml, HOL4, Coq, and Isabelle code.

- Participants: Francesco Zappa Nardelli, Peter Sewell and Scott Owens
- Contact: Francesco Zappa Nardelli
- URL: http://www.cl.cam.ac.uk/~pes20/lem/

## 5.6. Lucid Synchrone

FUNCTIONAL DESCRIPTION: Lucid Synchrone is a language for the implementation of reactive systems. It is based on the synchronous model of time as provided by Lustre combined with features from ML languages. It provides powerful extensions such as type and clock inference, type-based causality and initialization analysis and allows to arbitrarily mix data-flow systems and hierarchical automata or flows and valued signals.

RELEASE FUNCTIONAL DESCRIPTION: The language is still used for teaching and in our research but we do not develop it anymore. Nonetheless, we have integrated several features from Lucid Synchrone in new research prototypes described below. The Heptagon language and compiler are a direct descendent of it. The new language Zélus for hybrid systems modeling borrows many features originaly introduced in Lucid Synchrone.

- Contact: Marc Pouzet
- URL: http://www.di.ens.fr/~pouzet/lucid-synchrone/

## 5.7. Lucy-n

*Lucy-n: an n-synchronous data-flow programming language*
FUNCTIONAL DESCRIPTION: Lucy-n is a language to program in the n-synchronous model. The language is similar to Lustre with a buffer construct. The Lucy-n compiler ensures that programs can be executed in bounded memory and automatically computes buffer sizes. Hence this language allows to program Kahn networks, the compiler being able to statically compute bounds for all FIFOs in the program.

- Participants: Adrien Guatto, Albert Cohen, Louis Mandel and Marc Pouzet
- Contact: Albert Cohen
- URL: https://www.lri.fr/~mandel/lucy-n/

## 5.8. Ott

FUNCTIONAL DESCRIPTION: Ott is a tool for writing definitions of programming languages and calculi. It takes as input a definition of a language syntax and semantics, in a concise and readable ASCII notation that is close to what one would write in informal mathematics. It generates output:

a LaTeX source file that defines commands to build a typeset version of the definition,

a Coq version of the definition,

an Isabelle version of the definition, and

a HOL version of the definition.

Additionally, it can be run as a filter, taking a LaTeX/Coq/Isabelle/HOL source file with embedded (symbolic) terms of the defined language, parsing them and replacing them by typeset terms.

The main goal of the Ott tool is to support work on large programming language definitions, where the scale makes it hard to keep a definition internally consistent, and to keep a tight correspondence between a definition and implementations. We also wish to ease rapid prototyping work with smaller calculi, and to make it easier to exchange definitions and definition fragments between groups. The theorem-prover backends should enable a smooth transition between use of informal and formal mathematics.

- Participants: Francesco Zappa Nardelli, Peter Sewell and Scott Owens
- Contact: Francesco Zappa Nardelli
- URL: http://www.cl.cam.ac.uk/~pes20/ott/

## 5.9. PPCG

FUNCTIONAL DESCRIPTION: PPCG is our source-to-source research tool for automatic parallelization in the polyhedral model. It serves as a test bed for many compilation algorithms and heuristics published by our group, and is currently the best automatic parallelizer for CUDA and OpenCL (on the Polybench suite).

- Participants: Albert Cohen, Riyadh Baghdadi, Sven Verdoolaege and Tobias Grosser
- Contact: Sven Verdoolaege
- URL: http://freshmeat.net/projects/ppcg

## 5.10. ReactiveML

FUNCTIONAL DESCRIPTION: ReactiveML is a programming language dedicated to the implementation of interactive systems as found in graphical user interfaces, video games or simulation problems. ReactiveML is based on the synchronous reactive model due to Boussinot, embedded in an ML language (OCaml).

The Synchronous reactive model provides synchronous parallel composition and dynamic features like the dynamic creation of processes. In ReactiveML, the reactive model is integrated at the language level (not as a library) which leads to a safer and a more natural programming paradigm.

- Participants: Cédric Pasteur, Guillaume Baudart and Louis Mandel
- Contact: Guillaume Baudart

## 5.11. SundialsML

*Sundials/ML*
KEYWORDS: Simulation - Mathematics - Numerical simulations
SCIENTIFIC DESCRIPTION: Sundials/ML is a comprehensive OCaml interface to the Sundials suite of numerical solvers (CVODE, CVODES, IDA, IDAS, KINSOL). Its structure mostly follows that of the Sundials library, both for ease of reading the existing documentation and for adapting existing source code, but several changes have been made for programming convenience and to increase safety, namely:

solver sessions are mostly configured via algebraic data types rather than multiple function calls,

errors are signalled by exceptions not return codes (also from user-supplied callback routines),

user data is shared between callback routines via closures (partial applications of functions),

vectors are checked for compatibility (using a combination of static and dynamic checks), and

explicit free commands are not necessary since OCaml is a garbage-collected language.
FUNCTIONAL DESCRIPTION: Sundials/ML is a comprehensive OCaml interface to the Sundials suite of numerical solvers (CVODE, CVODES, IDA, IDAS, KINSOL, ARKODE).

RELEASE FUNCTIONAL DESCRIPTION: Adds support for v3.1.x of the Sundials Suite of numerical solvers.

Notably this release adds support for the new generic matrix and linear solver interfaces. The OCaml interface changes but the library is backward compatible with Sundials 2.7.0.

OCaml 4.02.3 or greater is now required and optionally OCamlMPI 1.03.

* New Sundials.Matrix and Sundials.LinearSolver modules. * Better treatment of integer type used for matrix indexing. * Refactor Dls and Sls modules into Sundials.Matrix. * Add confidence intervals to performance graph. * Miscellaneous improvements to configure script. * Potential incompatibility: changes to some label names: com_fn -> comm, ite_type -> iter. * Untangle the ARKODE mass-solver interface from the Jacobian interface.

- Participants: Jun Inoue, Marc Pouzet and Timothy Bourke
- Partner: UPMC
- Contact: Marc Pouzet
- URL: http://inria-parkas.github.io/sundialsml/

## 5.12. Zelus

SCIENTIFIC DESCRIPTION: The Zélus implementation has two main parts: a compiler that transforms Zélus programs into OCaml programs and a runtime library that orchestrates compiled programs and numeric solvers. The runtime can use the Sundials numeric solver, or custom implementations of well-known algorithms for numerically approximating continuous dynamics.
FUNCTIONAL DESCRIPTION: Zélus is a new programming language for hybrid system modeling. It is based on a synchronous language but extends it with Ordinary Differential Equations (ODEs) to model continuous-time behaviors. It allows for combining arbitrarily data-flow equations, hierarchical automata and ODEs. The language keeps all the fundamental features of synchronous languages: the compiler statically ensure the absence of deadlocks and critical races, it is able to generate statically scheduled code running in bounded time and space and a type-system is used to distinguish discrete and logical-time signals from continuous-time ones. The ability to combines those features with ODEs made the language usable both for programming discrete controllers and their physical environment.

- Participants: Marc Pouzet and Timothy Bourke
- Contact: Marc Pouzet

# 6. New Results

## 6.1. Verified compilation of Lustre

**Participants:** Timothy Bourke, Lélio Brun, Marc Pouzet.

Synchronous dataflow languages and their compilers are increasingly used to develop safety-critical applications, like fly-by-wire controllers in aircraft and monitoring software for power plants. A striking example is the SCADE Suite tool of ANSYS/Esterel Technologies which is DO-178B/C qualified for the aerospace and defense industries. This tool allows engineers to develop and validate systems at the level of abstract block diagrams that are automatically compiled into executable code.

Formal modeling and verification in an interactive theorem prover can potentially complement the industrial certification of such tools to give very precise definitions of language features and increased confidence in their correct compilation; ideally, right down to the binary code that actually executes.

This year we continued work on our verified Lustre compiler. We developed a new semantic model for the modular reset feature provided by the Scade language and required for the compilation of hierarchical state machines. This work was presented at the SCOPES workshop in Germany in May [17]. Work continues on connecting this semantic model to the intermediate compilation target.

We completed work on generalizing the compiler to treat clocked arguments. This involved changes to our intermediate Obc language and the addition of a pass to add some (necessary) variable initializations in an efficient way. This work was accepted for presentation at the Journées Francophones des Langages Applicatifs in 2019.

## 6.2. Julia Subtyping Reconstructed

**Participant:** Francesco Zappa Nardelli.

Julia is a programming language recently designed at MIT to support the needs of the scientific community. Julia occupies a unique position in the design landscape, it is a dynamic language with no type system, yet it has a surprisingly rich set of types and type annotations used to specify multimethod dispatch. The types that can be expressed in function signatures include parametric union types, covariant tuple types, parametric user-defined types with single inheritance, invariant type application, and finally types and values can be reified to appear in signatures. With Vitek started a research project to study the design and the pragmatic use of the Julia language. At first we focused on the Julia subtyping algorithm. We studied the empirical evidence that users appeal to all the features provided by Julia and we report on a formalisation and implementation of the subtyping algorithm. This has been published in [15]. We are pursuing this line of research studying of the algorithm advances of Julia can be integrated into other programming languages.

## 6.3. Comparing Designs for Gradual Types

**Participant:** Francesco Zappa Nardelli.

The enduring popularity of dynamically typed languages has given rise to a cottage industry of static type systems, often called gradual type systems, that let developers annotate legacy code piecemeal. Type soundness for a program which mixes typed and untyped code does not ensure the absence of errors at runtime, rather it means that some errors will caught at type checking time, while other will be caught as the program executes. After a decade of research it is clear that the combination of mutable state, self references and subtyping presents interesting challenges to designers of gradual type systems. We have reviewed the state of the art in gradual typing for objects, and introduced a class-based object calculus with a static type system, dynamic method dispatch, transparent wrappers and dynamic class generation that we use to model key features of several gradual type systems by translation to it, and discuss the implications of the respective designs. This has been published in [18].

## 6.4. Fast and reliable unwinding via DWARF tables

**Participants:** Theophile Bastian, Francesco Zappa Nardelli.

DWARF is a widely-used debugging data format. DWARF is obviously relied upon by debuggers, but it plays an unexpected role in the runtime of high-level programming languages and in the implementation of program analysis tools. The debug information itself can be pervaded by subtle bugs, making the whole infrastructure unreliable. In this project we are investigating techniques and tools to perform validation and synthesis of the DWARF stack unwinding tables, to speedup DWARF-based unwinding, as well as exploring adventurous projects that can be built on top of reliable DWARF information.

At the time of writing, we have a tool that can validate DWARF unwind tables generated by mainstream compilers; the approach is effective, we found a problem in Clang table generation and several in GLIBC inline-assembly snippets. We also designed and implemented a tool that can synthesise DWARF unwind tables from binary that lacks them (e.g. because the compiler did not generate them - immediate applications: JITs assembly, inline assembly, ...). Additionally we have designed and implemented a ahead-of-time compiler of DWARF unwind tables to assembly, and an ad-hoc unwinder integrated with the defacto standard unwinder libuwind. It can speed up unwinding by a factor between 25x and 60x (depending on application), with a 2.5x size overhead for unwind information.

Discussion is in progress to get these tools included in mainstream tool (e.g. the GNU profiler Perf).

## 6.5. Sundials/ML: OCaml interface to Sundials Numeric Solvers

**Participants:** Timothy Bourke, Marc Pouzet.

This year we made major updates to the Sundials/ML OCaml interface to support v3.1.x of the Sundials Suite of numerical solvers.

This release adds support for the new generic matrix and linear solver interfaces. Major work was required to add these new modules, update the existing solver interfaces, and ensure backwards compatibility with Sundials to v2.7.0 (which is still the version installed by Debian stable). We also improved our treatment of integer types used in indexing, refactored the Dls and Sls matrix modules, improved our generation of performance stats (by adding confidence intervals), made the configure script more robust, and untangle the mass-solver and Jacobian interfaces of the ARKODE solver.

## 6.6. Zélus

**Participants:** Timothy Bourke, Marc Pouzet.

This year, we made a major revision of the language and compiler, called now the version 2. The language now deal with higher order functions. All the static analyses, type inference, causality inference and the initialization analysis has been extended. The code generation has also been improved, in particular the interface with the numeric solver. Several larger examples have been written.

A paper that present the overall approach followed in ZELUS has been published [12].

## 6.7. Deterministic Concurrency: A Clock-Synchronised Shared Memory Approach

**Participant:** Marc Pouzet.

Synchronous programming (SP) provides deterministic concurrency. So far, however, communication has been constrained to a set of primitive clock-synchronised shared memory (scm) data types, such as data-flow registers, streams and signals with restricted read and write accesses that limit modularity and behavioural abstractions. In the paper [23], we propose an extension to the SP theory which retains the advantages of deterministic concurrency, but allows communication to occur at higher levels of abstraction than currently supported by SP data types. Our approach is as follows. To avoid data races, each csm type publishes a policy interface for specifying the admissibility and precedence of its access methods. Each instance of the csm type has to be policy-coherent, meaning it must behave deterministically under its own policy—a natural requirement if the goal is to build deterministic sys- tems that use these types. In a policy-constructive system, all access methods can be scheduled in a policy-conformant way for all the types without deadlocking. In this paper, we show that a policy-constructive program exhibits deterministic concurrency in the sense that all policy-conformant interleavings produce the same input-output behaviour. Policies are conservative and support the csm types existing in current SP languages. This work is a follower of a old work we did in 2009, published at LCTES about scheduling policies.

## 6.8. Compiling synchronous languages for multi-processor implementations

**Participants:** Guillaume Iooss, Albert Cohen, Timothy Bourke, Marc Pouzet.

This work was performed with industrial partners in the context of the ASSUME project.

We have continued to improve our front-end tools for a use case provided by Airbus. This tool now generates three kinds of monolithic Lustre program, which are taken as an input of the Lopht tool (AOSTE team), which in turn generates an executable for the Kalray MPPA. In particular, one of the code generators is based on the hyper-period-expansion transformation, which unrolls the computation and generates a single step function running at the slowest period. This transformation allows managing the multi-periodic aspect of the application at the source level. Together with the work of the AOSTE team and Airbus, it allows us to execute the full application on a MPPA (TRL-5 Airbus certification level).

We have also improved the front-end tools for the use case provided by Safran. These tools ware integrated into the Heptagon compiler. Many improvements to the parser and many convenient program transformations (tuple and array destruction, equation clustering, ...) were implemented in the Heptagon compiler in order to treat this use case and enable the Lopht tool to extract the best performance. In particular, we have investigated the impact of inlining on the degree of parallelism exposed by the use-case application.

In addition to the work described above, we have defined a language extension for 1-synchronous clocks, strictly periodic clocks with a single activation. We show that we can derive a scheduling problem from the clock constraints in a program. However, solving these constraints by using an interesting cost functions (such as WCET load balancing across the different phases of a period) with an ILP does not scale for the two use cases. Thus, we used the fact that we do not need the optimal solution to fall back on heuristics, which finds a good solution within acceptable bounds. We have also investigated the effect of a non-determinism operator on the scheduling constraints, which gives extra freedom for choosing a schedule.

In collaboration (this year) with Dumitru Potop-Butucaru and Keryan Didier (Inria, AOSTE team); Jean Souyris and Vincent Bregeon (Airbus); Philippe Baufreton and Jean-Marie Courtelle (Safran).

In collaboration with ANSYS, a compilation technique has been designed for compiling SCADE to multi-core [24].

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Bilateral Contracts with Industry

Polly Labs contract with ARM, 2015-2019, with the participation of Qualcomm, Xilinx and Facebook (human resources, consulting services and and hiring former PARKAS members).

## 7.2. Bilateral Grants with Industry

In 2018 Francesco Zappa Nardelli was awarded a Google Research Fellowship to pursue the work on DWARF unwinding, about 50k euros.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

The Inria Project Lab (IPL) *Modeliscale* treats the modelling and analysis of Cyber-Physical Systems at large scale. The PARKAS team contributes their expertise in programming language design for reactive and hybrid systems to this multi-team effort.

### *8.1.1. ANR*

ANR/CHIST-ERA DIVIDEND project, 2013-2018.

### *8.1.2. FUI: Fonds unique interministériel*

Modeliscale contract (AAP-24). Using Modelica at scale to model and simulate very large Cyber-Physical Systems. Principal industrial partner: Dassault-Systèmes. Inria contacts are Benoit Caillaud (HYCOMES, Rennes) and Marc Pouzet (PARKAS, Paris).

### *8.1.3. Others*

Marc Pouzet is scientific advisor for the Esterel-Technologies/ANSYS company.

## 8.2. European Initiatives

### *8.2.1. H2020 Projects*

Program: H2020 "Smart Anything Everywhere (SAE)" initiative

Project acronym: TETRAMAX

Project title: Technology Transfer via Multinational Application Experiments

Duration: September 2017 – August 2021

Coordinator: Rainer Leupers

Other partners: Rheinisch-Westfaelische Technische Hochschule Aachen, RWTH (Germany); AMG Technology Ood, AMGT (Bulgaria); Ruhr-Universitaet Bochum, RUB (Germany); Budapesti Muszaki Es Gazdasagtudomanyi Egyetem, BME (Hungary); Universitat Politecnica De Catalunya, UPC (Spain); Control Data Systems Srl, CDS (Romania); Chalmers Tekniska Hoegskola Ab, CHALMERS, (Sweden); Technische Universiteit Delft, TuDelft (Netherlands); The University Of Edinburgh, UEDIN, (United Kingdom); Fundingbox Accelerator Sp z o.o., FBOX, (Poland); Universiteit Gent, UGENT (Belgium); Vysoka Skola Banska -Technicka Univerzita Ostrava, IT4I, (Czech Republic); Institut Jozef Stefan, JSI, Slovenia, Techmo Spolka z o.o., TECHMO (Poland); Universita Di Pisa, PISA (Italy); Tallinna Tehnikaulikool, TTU (Estonia); Tty-Saatio,TUT (Finland); Think Silicon Ereyna Kai Technologia Anonymi, Etairia, THINKS (Greece); Technische Universitaet Muenchen, TUM (Germany); Sveuciliste U Zagrebu Fakultet Elektrotehnike I Racunarstva, UZAGREB, (Croatia); Zentrum Fur Innovation Und Technik In Nordrhein-Westfalen GmbH, ZENIT (Germany).

Abstract: The overall ambition of TETRAMAX is building and leveraging a European Competence Center Network in customized low-energy computing, providing easy access for SMEs and mid-caps to novel CLEC technologies via local contact points. This is a bidirectional interaction: SMEs can demand CLEC technologies and solutions via the network, and vice versa academic research institutions can actively and effectively offer their new technologies to European industries. Furthermore, TETRAMAX wants to support 50+ industry clients and 3rd parties with innovative technologies, using different kinds of Technology Transfer Experiments (TTX) to accelerate innovation within European industries and to create a competitive advantage in the global economy.

### *8.2.2. Collaborations in European Programs, Except FP7 & H2020*

Program: ITEA3

Project acronym: 14014 ASSUME

Project title: Affordable Safe & Secure Mobility Evolution

Duration: September 2015 – December 2018

Coordinator: Dumitru Potop Butucaru

Other partners: *France:* Airbus, École Normale Supérieure (ENS), Esterel Technologies, Kalray SA, Safran Aircraft Engines SAS SNECMA, Safran Electronics & Defense Sagem, Sorbonne Université, Thales; *Germany:* AbsInt Angewandte Informatik GmbH, Assystem Germany GmbH, BTC Embedded Systems AG, Daimler AG, FZI Forschungszentrum Informatik, Karlsruhe Institute of Technology (KIT), Kiel University, Model Engineering Solutions GmbH, OFFIS, Robert Bosch GmbH, Technical University of Munich; *Netherlands:* Eindhoven University of Technology, NXP Semiconductors Netherlands BV, Recore Systems BV, TNO, University of Twente, VDL Enabling Transport Solutions, Verum Software Tools BV; *Sweden:* Arcticus Systems AB, FindOut Technologies AB, KTH (Royal Institute of Technology), Mälardalen University, Scania; *Turkey:* Arçelik, Ericsson Ar-Ge, Ford Otosan, Havelsan, KoçSistem, UNIT Information Technologies R&D Ltd.

Abstract: Future mobility solutions will increasingly rely on smart components that continuously monitor the environment and assume more and more responsibility for a convenient, safe and reliable operation. Currently the single most important roadblock for this market is the ability to come up with an affordable, safe multi-core development methodology that allows industry to deliver trustworthy new functions at competitive prices. ASSUME will provide a seamless engineering methodology, which addresses this roadblock on the constructive and analytic side.

## 8.3. International Initiatives

### *8.3.1. Inria Associate Teams Not Involved in an Inria International Labs*

#### *8.3.1.1. POLYFLOW*

Title: Polyhedral Compilation for Data-Flow Programming Languages

International Partner (Institution - Laboratory - Researcher):

IISc Bangalore (India) - Department of Computer Science and Automation (CSA) - Uday Kumar Reddy Bondhugula

Start year: 2016

See also: http://polyflow.gforge.inria.fr

The objective of the associate team is to foster collaborations on fundamental and applied research. It also supports training sessions, exchange of undergraduate and master students, and highlighting opportunities in the partners' research, education and economic environments.

Polyhedral techniques for program transformation are now used in several proprietary and open source compilers. However, most of the research on polyhedral compilation has focused on imperative languages, where computation is specified in terms of computational statements within nested loops and control structures. Graphical data-flow languages, where there is no notion of statements or a schedule specifying their relative execution order, have so far not been studied using a powerful transformation or optimization approach. These languages are extremely popular in the system analysis, modeling and design of embedded reactive control applications. They also underline the construction of domain-specific languages and compiler intermediate representations. The execution semantics of data-flow languages impose a different set of challenges for compilation and optimization. We are studying techniques enabling the extraction of a polyhedral representation from data-flow programs, to transform them with the goal of generating memory-efficient and high-performance code for modern architectures.

The research conducted in PolyFlow covers both fundamental and applied aspects. The partners also emphasize the development of solid research tools. The associate team will facilitate their dissemination as free software and their exploitation through industrial collaborations.

### *8.3.2. Participation in Other International Programs*

- VerticA (Francesco Zappa Nardelli), 2017-2020, joint project with Northeastern University, USA, financed by the ONR (Office of Naval Research), $1.5M (subcontract for $150k).

*8.3.2.1. Indo-French Center of Applied Mathematics*

**POLYFLOW**

Title: Polyhedral Compilation for Data-Flow Programming Languages

International Partner (Institution - Laboratory - Researcher):

IISc Bangalore (India) - Uday Kumar Reddy Bondhugula

Duration: 2016 - 2018

Start year: 2016

The objective of the associate team is to foster collaborations on fundamental and applied research. It also supports training sessions, exchange of undergraduate and master students, and highlighting opportunities in the partners' research, education and economic environments. Polyhedral techniques for program transformation are now used in several proprietary and open source compilers. However, most of the research on polyhedral compilation has focused on imperative languages, where computation is specified in terms of computational statements within nested loops and control structures. Graphical data-flow languages, where there is no notion of statements or a schedule specifying their relative execution order, have so far not been studied using a powerful transformation or optimization approach. These languages are extremely popular in the system analysis, modeling and design of embedded reactive control applications. They also underline the construction of domain-specific languages and compiler intermediate representations. The execution semantics of data-flow languages impose a different set of challenges for compilation and optimization. We are studying techniques enabling the extraction of a polyhedral representation from data-flow programs, to transform them with the goal of generating memory-efficient and high-performance code for modern architectures. The research conducted in PolyFlow covers both fundamental and applied aspects. The partners also emphasize the development of solid research tools. The associate team will facilitate their dissemination as free software and their exploitation through industrial collaborations.

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Selection*

*9.1.1.1. Member of the Conference Program Committees*

- T. Bourke served on the PC of the International Conference on Embedded Software (EMSOFT 2018).
- T. Bourke served on the PC for the Journées Francophones des Langages Applicatifs (JFLA 2018).
- T. Bourke served on the PC of the American Modelica Conference 2018.
- T. Bourke served on the PC of the Japanese Modelica Conference 2018.
- T. Bourke served on the PC of the International Workshop on Software and Compilers for Embedded Systems (SCOPES 2018).
- F. Zappa Nardelli will serve on the PC of OOPSLA 2019 (International Conference on Object-Oriented Programming, Systems, Languages & Applications).
- M. Pouzet served on the PC of the International Workshop on Software and Compilers for Embedded Systems (SCOPES 2018).
- M. Pouzet served on the PC of the International Conference on Embedded Software (EMSOFT 2018).
- M. Pouzet served on the PC of the International Conference on Principles and Practice of Declarative Programming (PPDP 2018).

- M. Pouzet served on the PC of the International Forum on specification & Design Languages (FDL 2018).

*9.1.1.2. Reviewer*

- T. Bourke reviewed for the International Joint Conference on Automated Reasoning (IJCAR 2018)
- T. Bourke reviewed for the International Conference on Interactive Theorem Proving (ITP 2018)
- T. Bourke reviewed for the International Symposium on Principles and Practice of Declarative Programming (PPDP 2018)

### *9.1.2. Journal*

*9.1.2.1. Reviewer - Reviewing Activities*

- T. Bourke was a reviewer for Science of Computer Programming.

### *9.1.3. Invited Talks*

- T. Bourke was invited to present a seminar at the Collège de France in January 2018: *La vérification formelle d'un compilateur Lustre*.
- T. Bourke was invited to present an LSV Seminar (Laboratoire Spécification et Vérification) at the ENS Cachan in January 2018: *Compiling a Synchronous Language with Timers for Simulation*
- T. Bourke was invited to present to the Société Informatique de France doctoral seminar at the École normale supérieure in June 2018: *10 lignes de logique pour 1 ligne de code (correct)*
- F. Zappa Nardelli was invited to present to the DeepSpec Workshop in July 2018 on *Debugging Debug Information.*
- T. Bourke and M. Pouzet participated by invitation in the Shonan Meeting 136 on *Functional Stream Libraries and Fusion: What's next?*, in Japan in October 2018.
- M. Pouzet was invited to give a lecture at the International summer school at Marktoberdorf, in July 2018.

### *9.1.4. Research Administration*

- F. Zappa Nardelli will chair the part-time assistant professor recruitment committee at École Polytechinque.

## 9.2. Teaching - Supervision - Juries

### *9.2.1. Teaching*

Master: F. Zappa Nardelli: "A Programmer's introduction to Computer Architectures and Operating Systems" (M1), 45h, École Polytechnique, France

Master: A. Cohen & F. Zappa Nardelli, "Semantics, languages and algorithms for multicore programming", Lecture, 12h+9h, M2, MPRI: Ecole normale supeérieure and Université Paris Diderot, France

Master : M. Pouzet & T. Bourke: "Synchronous Systems" (M2), Lectures and TDs, MPRI, France

Master: T. Bourke participated in reviewing the L3 and M1 internships of students at the ENS, France.

Licence : M. Pouzet & T. Bourke: "Operating Systems" (L3), Lectures and TDs, ENS, France.

Licence : T. Bourke, "Digital Systems" (L3), Lectures and TDs, ENS, France

Marc Pouzet is Director of Studies for the CS department, at ENS.

### *9.2.2. Supervision*

PhD in progress : Lélio Brun, 3rd year, supervised by T. Bourke and M. Pouzet.

PhD in progress : Chandan Reddy, 3rd year, supervised by A. Cohen.

PhD : Jie Zhao, 3rd year, supervised by A. Cohen, defended in December 2018.

PhD in progress : Basile Clément, 1st year, supervised by F. Zappa Nardelli and A. Cohen.

#### *9.2.3. Juries*

- Francesco Zappa Nardelli was jury member of the PhD thesis of Francois Ginraud, Grenoble, Jan 2018.
- T. Bourke was an examiner for the thesis of Jiangchao LIU at the École normale supérieure in February 2018.
- T. Bourke was an examiner for the thesis of Hai NGUYEN VAN at the Université Paris-Sud in September 2018.
- T. Bourke was an examiner for the thesis of Narjes JOMAA at the Université de Lille in December 2018.
- F. Zappa Nardelli was an examiner for the thesis of Jie Zhao at the École normale supérieure in December 2018.

### 9.3. Popularization

#### *9.3.1. Internal or external Inria responsibilities*

- F. Zappa Nardelli is member of the CES of Inria.

# 10. Bibliography

## Major publications by the team in recent years

[1] T. BOURKE, L. BRUN, P.-E. DAGAND, X. LEROY, M. POUZET, L. RIEG.*A Formally Verified Compiler for Lustre*, in "PLDI 2017 - 38th ACM SIGPLAN Conference on Programming Language Design and Implementation", Barcelone, Spain, ACM, June 2017, https://hal.inria.fr/hal-01512286

[2] T. BOURKE, F. CARCENAC, J.-L. COLAÇO, B. PAGANO, C. PASTEUR, M. POUZET.*A Synchronous Look at the Simulink Standard Library*, in "EMSOFT 2017 - 17th International Conference on Embedded Software", Seoul, South Korea, ACM Press, October 2017, 23, https://hal.inria.fr/hal-01575631

[3] T. BOURKE, J.-L. COLAÇO, B. PAGANO, C. PASTEUR, M. POUZET.*A Synchronous-based Code Generator For Explicit Hybrid Systems Languages*, in "International Conference on Compiler Construction (CC)", London, United Kingdom, LNCS, July 2015, https://hal.inria.fr/hal-01242732

[4] L. GÉRARD, A. GUATTO, C. PASTEUR, M. POUZET.*A modular memory optimization for synchronous data-flow languages: application to arrays in a lustre compiler*, in "Proceedings of the 13th ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, Tools and Theory for Embedded Systems", Beijing, China, ACM, June 2012, p. 51–60 [*DOI :* 10.1145/2248418.2248426], https://hal.inria.fr/hal-00728527

[5] J. C. JUEGA, S. VERDOOLAEGE, A. COHEN, J. I. GÓMEZ, C. TENLLADO, F. CATTHOOR.*Patterns for parallel programming on GPUs*, in "Patterns for parallel programming on GPUs", F. MAGOULÈS (editor), Saxe-Cobourg, 2013, vol. Evaluation of State-of-the-Art Parallelizing Compilers Generating CUDA Code for Heterogeneous CPU/GPU Computing, ISBN 978-1-874672-57-9, https://hal.archives-ouvertes.fr/hal-01257261

[6] L. MANDEL, F. PLATEAU, M. POUZET.*Static Scheduling of Latency Insensitive Designs with Lucy-n*, in "FMCAD 2011 - Formal Methods in Computer Aided Design", Austin, TX, United States, October 2011, https://hal.inria.fr/hal-00654843

[7] R. MORISSET, P. PAWAN, F. ZAPPA NARDELLI.*Compiler testing via a theory of sound optimisations in the C11/C++11 memory model*, in "PLDI 2013 - 34th ACM SIGPLAN conference on Programming language design and implementation", Seattle, WA, United States, ACM, June 2013, p. 187-196 [*DOI :* 10.1145/2491956.2491967], https://hal.inria.fr/hal-00909083

[8] A. POP, A. COHEN.*OpenStream: Expressiveness and Data-Flow Compilation of OpenMP Streaming Programs*, in "ACM Transactions on Architecture and Code Optimization", 2013, vol. 9, n$^{o}$ 4, Selected for presentation at the HiPEAC 2013 Conference [*DOI :* 10.1145/2400682.2400712], https://hal.inria.fr/hal-00786675

[9] J. SEVCIK, V. VAFEIADIS, F. ZAPPA NARDELLI, S. JAGANNATHAN, P. SEWELL.*CompCertTSO: A Verified Compiler for Relaxed-Memory Concurrency*, in "Journal of the ACM (JACM)", 2013, vol. 60, n$^{o}$ 3, p. art. 22:1-50 [*DOI :* 10.1145/2487241.2487248], https://hal.inria.fr/hal-00909076

[10] V. VAFEIADIS, T. BALABONSKI, S. CHAKRABORTY, R. MORISSET, F. ZAPPA NARDELLI.*Common compiler optimisations are invalid in the C11 memory model and what we can do about it*, in "POPL 2015 - 42nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages", Mumbai, India, January 2015, https://hal.inria.fr/hal-01089047

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] J. ZHAO.*A Combined Language and Polyhedral Approach for Heterogeneous Parallelism*, PSL Research University, December 2018, https://tel.archives-ouvertes.fr/tel-01988073

### Articles in International Peer-Reviewed Journal

[12] A. BENVENISTE, T. BOURKE, B. CAILLAUD, J.-L. COLAÇO, C. PASTEUR, M. POUZET.*Building a Hybrid Systems Modeler on Synchronous Languages Principles*, in "Proceedings of the IEEE", September 2018, vol. 106, n$^{o}$ 9, p. 1568 - 1592 [*DOI :* 10.1109/JPROC.2018.2858016], https://hal.inria.fr/hal-01879026

[13] T. BOURKE, J. INOUE, M. POUZET.*Sundials/ML: Connecting OCaml to the Sundials Numeric Solvers*, in "Electronic Proceedings in Theoretical Computer Science", December 2018, vol. 285, p. 101-130, https://arxiv.org/abs/1812.11668 [*DOI :* 10.4204/EPTCS.285.4], https://hal.inria.fr/hal-01967659

[14] S. RAMAKRISHNAN, A. GOHLKE, F. LI, J. COLEMAN, W. XU, J. E. ROTHMAN, F. PINCET.*High-Throughput Monitoring of Single Vesicle Fusion Using Freestanding Membranes and Automated Analysis*, in "Langmuir", May 2018, vol. 34, n$^{o}$ 20, p. 5849-5859 [*DOI :* 10.1021/ACS.LANGMUIR.8B00116], https://hal.sorbonne-universite.fr/hal-01954048

[15] F. ZAPPA NARDELLI, J. BELYAKOVA, A. PELENITSYN, B. CHUNG, J. BEZANSON, J. VITEK.*Julia Subtyping: A Rational Reconstruction*, in "Proceedings of the ACM on Programming Languages", 2018, vol. 27, OOPSLA, Article 113 [*DOI :* 10.1145/3276483], https://hal.inria.fr/hal-01882137

[16] 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° 1, p. 1 - 25 [*DOI :* 10.1145/3177961], https://hal.inria.fr/hal-01744426

### International Conferences with Proceedings

[17] T. BOURKE, L. BRUN, M. POUZET.*Towards a verified Lustre compiler with modular reset*, in "21st International Workshop on Software and Compilers for Embedded Systems (SCOPES 2018)", Sankt Goar, Germany, Proceedings of the 21st International Workshop on Software and Compilers for Embedded Systems (SCOPES 2018), ACM Press, May 2018, 4 [*DOI :* 10.1145/3207719.3207732], https://hal.inria.fr/hal-01817949

[18] B. CHUNG, P. LI, F. ZAPPA NARDELLI, J. VITEK.*KafKa: Gradual Typing for Objects*, in "ECOOP 2018 - 2018 European Conference on Object-Oriented Programming", Amsterdam, Netherlands, July 2018, https://hal.inria.fr/hal-01882148

[19] P. FEAUTRIER, A. COHEN, A. DARTE.*On polynomial Code Generation*, in "IMPACT 2018", Manchester, United Kingdom, January 2018, https://hal.inria.fr/hal-01958096

[20] J. SOUYRIS, K. DIDIER, D. POTOP-BUTUCARU, G. IOOSS, T. BOURKE, A. COHEN, M. POUZET.*Automatic Parallelization from Lustre Models in Avionics*, in "ERTS2 2018 - 9th European Congress Embedded Real-Time Software and Systems", Toulouse, France, Proceedings of the 9th European Congress on Embedded Real-Time Software and Systems (ERTS2), 3AF - Association Aéronautique Astronautique de France and SEE - Société de l'électricité, de l'électronique et des technologies de l'information et de la communication and SIA - Société de Ingénieurs de l'Automobile, January 2018, p. 1-4, https://hal.inria.fr/hal-01714054

[21] J. ZHAO, M. KRUSE, A. COHEN.*A polyhedral compilation framework for loops with dynamic data-dependent bounds*, in "CC'18 - 27th International Conference on Compiler Construction", Vienna, Austria, ACM Press, February 2018 [*DOI :* 10.1145/3178372.3179509], https://hal.inria.fr/hal-01720368

[22] O. ZINENKO, S. VERDOOLAEGE, C. REDDY, J. SHIRAKO, T. GROSSER, V. SARKAR, A. COHEN.*Modeling the conflicting demands of parallelism and Temporal/Spatial locality in affine scheduling*, in "CC'18 - 27th International Conference on Compiler Construction", Vienna, Austria, ACM Press, February 2018 [*DOI :* 10.1145/3178372.3179507], https://hal.inria.fr/hal-01751823

### Conferences without Proceedings

[23] J. AGUADO, M. MENDLER, M. POUZET, P. ROOP, R. VON HANXLEDEN.*Deterministic Concurrency: A Clock-Synchronised Shared Memory Approach*, in "ESOP 2018 - European Symposium on Programming", Thessaloniki, Greece, April 2018, https://hal.archives-ouvertes.fr/hal-01960404

[24] J.-L. COLAÇO, B. PAGANO, C. PASTEUR, M. POUZET.*Scade 6: from a Kahn Semantics to a Kahn Implementation for Multicore*, in "Forum on specification & Design Languages (FDL)", Munich, Germany, September 2018, https://hal.archives-ouvertes.fr/hal-01960410

### Research Reports

[25] K. DIDIER, D. POTOP-BUTUCARU, G. IOOSS, A. COHEN, J. SOUYRIS, P. BAUFRETON, A. GRAILLAT.*Parallelisation efficace de larges applications temps-reel*, Inria Paris, June 2018, n° RR-9180, https://hal.inria.fr/hal-01810176

[26] O. ZINENKO, L. CHELINI, T. GROSSER.*Declarative Transformations in the Polyhedral Model*, Inria ; ENS Paris - Ecole Normale Supérieure de Paris ; ETH Zurich ; TU Delft ; IBM Zürich, December 2018, n$^{o}$ RR-9243, https://hal.inria.fr/hal-01965599

## References in notes

[27] G. BAUDART.*A synchronous approach to quasi-periodic systems*, PSL Research University, March 2017, https://tel.archives-ouvertes.fr/tel-01507595

# Project-Team PI.R2

## Design, study and implementation of languages for proofs and programs

IN COLLABORATION WITH: Institut de Recherche en Informatique Fondamentale

IN PARTNERSHIP WITH:
**CNRS**

**Université Denis Diderot (Paris 7)**

RESEARCH CENTER
**Paris**

THEME
**Proofs and Verification**

# Table of contents

## Project-Team PI.R2

*Creation of the Team: 2009 January 01, updated into Project-Team: 2011 January 01*

**Keywords:**

**Computer Science and Digital Science:**
A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.11. - Proof languages
A2.4.3. - Proofs
A7.2. - Logic in Computer Science
A8.1. - Discrete mathematics, combinatorics
A8.4. - Computer Algebra

**Other Research Topics and Application Domains:**
B6.1. - Software industry
B6.6. - Embedded systems

# 1. Team, Visitors, External Collaborators

**Research Scientists**
Thierry Coquand [University of Gothenburg, Senior Researcher]
Pierre-Louis Curien [Team leader, CNRS, Senior Researcher, HDR]
Yves Guiraud [Inria, Researcher]
Hugo Herbelin [Inria, Senior Researcher, HDR]
Jean-Jacques Lévy [Inria, Emeritus, HDR]
Alexis Saurin [CNRS, Researcher]
Matthieu Sozeau [Inria, Researcher]

**Faculty Members**
Pierre Letouzey [Univ Denis Diderot, Associate Professor]
Yann Régis-Gianas [Univ Denis Diderot, Associate Professor]

**External Collaborators**
Thibaut Girka [Univ Denis Diderot, until Sep 2018]
Jovana Obradović [Univ Denis Diderot]

**Technical Staff**
Daniel de Rauglaudre [Inria, from Jul 2015]
Thierry Martinez [Inria]

**PhD Students**
Antoine Allioux [Inria, from Sep 2018]
Cédric Ho Thanh [Univ Denis Diderot]
Cyprien Mangin [Ecole polytechnique, until Aug 2018]
Théo Zimmermann [Univ Denis Diderot]

**Post-Doctoral Fellows**
Eric Finster [Inria]
Kailiang Ji [Inria]
Exequiel Rivas Gadda [Inria]

**Visiting Scientist**

Ying Jiang [Chinese Academy of Sciences, from Nov 2018]

**Administrative Assistants**

Mathieu Mourey [Inria, from Sep 2018]

Sandrine Verges [Inria, until Jun 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

The research conducted in $\pi r^2$ is devoted both to the study of foundational aspects of formal proofs and programs and to the development of the Coq proof assistant software, with a focus on the dependently typed programming language aspects of Coq. The team acts as one of the strongest teams involved in the development of Coq as it hosts in particular the current coordinator of the Coq development team.

Since 2012, the team has also extended its scope to the study of the homotopy of rewriting systems, which shares foundational tools with recent advanced works on the semantics of type theories.

# 3. Research Program

## 3.1. Proof theory and the Curry-Howard correspondence

### *3.1.1. Proofs as programs*

Proof theory is the branch of logic devoted to the study of the structure of proofs. An essential contributor to this field is Gentzen [83] who developed in 1935 two logical formalisms that are now central to the study of proofs. These are the so-called "natural deduction", a syntax that is particularly well-suited to simulate the intuitive notion of reasoning, and the so-called "sequent calculus", a syntax with deep geometric properties that is particularly well-suited for proof automation.

Proof theory gained a remarkable importance in computer science when it became clear, after genuine observations first by Curry in 1958 [78], then by Howard and de Bruijn at the end of the 60's [95], [114], that proofs had the very same structure as programs: for instance, natural deduction proofs can be identified as typed programs of the ideal programming language known as $\lambda$-calculus.

This proofs-as-programs correspondence has been the starting point to a large spectrum of researches and results contributing to deeply connect logic and computer science. In particular, it is from this line of work that Coquand and Huet's Calculus of Constructions [75], [76] stemmed out – a formalism that is both a logic and a programming language and that is at the source of the Coq system [113].

### *3.1.2. Towards the calculus of constructions*

The $\lambda$-calculus, defined by Church [73], is a remarkably succinct model of computation that is defined via only three constructions (abstraction of a program with respect to one of its parameters, reference to such a parameter, application of a program to an argument) and one reduction rule (substitution of the formal parameter of a program by its effective argument). The $\lambda$-calculus, which is Turing-complete, i.e. which has the same expressiveness as a Turing machine (there is for instance an encoding of numbers as functions in $\lambda$-calculus), comes with two possible semantics referred to as call-by-name and call-by-value evaluations. Of these two semantics, the first one, which is the simplest to characterise, has been deeply studied in the last decades [66].

To explain the Curry-Howard correspondence, it is important to distinguish between intuitionistic and classical logic: following Brouwer at the beginning of the $20^{\text{th}}$ century, classical logic is a logic that accepts the use of reasoning by contradiction while intuitionistic logic proscribes it. Then, Howard's observation is that the proofs of the intuitionistic natural deduction formalism exactly coincide with programs in the (simply typed) $\lambda$-calculus.

A major achievement has been accomplished by Martin-Löf who designed in 1971 a formalism, referred to as modern type theory, that was both a logical system and a (typed) programming language [105].

In 1985, Coquand and Huet [75], [76] in the Formel team of Inria-Rocquencourt explored an alternative approach based on Girard-Reynolds' system $F$ [84], [109]. This formalism, called the Calculus of Constructions, served as logical foundation of the first implementation of Coq in 1984. Coq was called CoC at this time.

### *3.1.3. The Calculus of Inductive Constructions*

The first public release of CoC dates back to 1989. The same project-team developed the programming language Caml (nowadays called OCaml and coordinated by the Gallium team) that provided the expressive and powerful concept of algebraic data types (a paragon of it being the type of lists). In CoC, it was possible to simulate algebraic data types, but only through a not-so-natural not-so-convenient encoding.

In 1989, Coquand and Paulin [77] designed an extension of the Calculus of Constructions with a generalisation of algebraic types called inductive types, leading to the Calculus of Inductive Constructions (CIC) that started to serve as a new foundation for the Coq system. This new system, which got its current definitive name Coq, was released in 1991.

In practice, the Calculus of Inductive Constructions derives its strength from being both a logic powerful enough to formalise all common mathematics (as set theory is) and an expressive richly-typed functional programming language (like ML but with a richer type system, no effects and no non-terminating functions).

## 3.2. The development of Coq

During 1984-2012 period, about 40 persons have contributed to the development of Coq, out of which 7 persons have contributed to bring the system to the place it was six years ago. First Thierry Coquand through his foundational theoretical ideas, then Gérard Huet who developed the first prototypes with Thierry Coquand and who headed the Coq group until 1998, then Christine Paulin who was the main actor of the system based on the CIC and who headed the development group from 1998 to 2006. On the programming side, important steps were made by Chet Murthy who raised Coq from the prototypical state to a reasonably scalable system, Jean-Christophe Filliâtre who turned to concrete the concept of a small trustful certification kernel on which an arbitrary large system can be set up, Bruno Barras and Hugo Herbelin who, among other extensions, reorganised Coq on a new smoother and more uniform basis able to support a new round of extensions for the next decade.

The development started from the Formel team at Rocquencourt but, after Christine Paulin got a position in Lyon, it spread to École Normale Supérieure de Lyon. Then, the task force there globally moved to the University of Orsay when Christine Paulin got a new position there. On the Rocquencourt side, the part of Formel involved in ML moved to the Cristal team (now Gallium) and Formel got renamed into Coq. Gérard Huet left the team and Christine Paulin started to head a Coq team bilocalised at Rocquencourt and Orsay. Gilles Dowek became the head of the team which was renamed into LogiCal. Following Gilles Dowek who got a position at École Polytechnique, LogiCal moved to the new Inria Saclay research center. It then split again, giving birth to ProVal. At the same time, the Marelle team (formerly Lemme, formerly Croap) which has been a long partner of the Formel team, invested more and more energy in the formalisation of mathematics in Coq, while contributing importantly to the development of Coq, in particular for what regards user interfaces.

After various other spreadings resulting from where the wind pushed former PhD students, the development of Coq got multi-site with the development now realised mainly by employees of Inria, the CNAM, Paris 7 and MINES.

In the last six years, Hugo Herbelin and Matthieu Sozeau coordinated the development of the system, the official coordinator hat passed from Hugo to Matthieu in August 2016. The ecosystem and development model changed greatly during this period, with a move towards an entirely distributed development model, integrating contributions from all over the world. While the system had always been open-source, its development team was relatively small, well-knit and gathered regularly at Coq working groups, and many developments on Coq were still discussed only by the few interested experts.

The last years saw a big increase in opening the development to external scrutiny and contributions. This was supported by the "core" team which started moving development to the open GitHub platform (including since 2017 its bug-tracker [60] and wiki), made its development process public, starting to use public pull requests to track the work of developers, organising yearly hackatons/coding-sprints for the dissemination of expertise and developers & users meetings like the Coq Workshop and CoqPL, and, perhaps more anecdotally, retransmitting Coq working groups on a public YouTube channel.

This move was also supported by the hiring of Maxime Dénès in 2016 as an Inria research engineer (in Sophia-Antipolis), and the work of Matej Košík (2-year research engineer). Their work involved making the development process more predictable and streamlined and to provide a higher level of quality to the whole system. In September 2018, a second engineer, Vincent Laporte, was hired. Yves Bertot, Maxime Dénès and Vincent Laporte are developing the Coq consortium, which aims to become the incarnation of the global Coq community and to offer support for our users.

Today, the development of Coq involves participants from the Inria project-teams pi.r2 (Paris), Marelle (Sophia-Antipolis), Toccata (Saclay), Gallinette (Nantes), Gallium (Paris), and Camus (Strasboug), the LIX at École Polytechnique and the CRI Mines-ParisTech. Apart from those, active collaborators include members from MPI-Saarbrucken (D. Dreyer's group), KU Leuven (B. Jacobs group), MIT CSAIL (A. Chlipala's group, which hosted an Inria/MIT engineer, and N. Zeldovich's group), the Institute for Advanced Study in Princeton (from S. Awodey, T. Coquand and V. Voevodsky's Univalent Foundations program) and Intel (M. Soegtrop). The latest released version Coq 8.8.0 had 40 contributors (counted from the start of 8.8 development) and the upcoming Coq 8.9 has 54.

On top of the developer community, there is a much wider user community, as Coq is being used in many different fields. The Software Foundations series, authored by academics from the USA, along with the reference Coq'Art book by Bertot and Castéran [67], the more advanced Certified Programming with Dependent Types book by Chlipala [72] and the recent book on the Mathematical Components library by Mahboubi, Tassi et al. provide resources for gradually learning the tool.

In the programming languages community, Coq is being taught in two summer schools, OPLSS and the DeepSpec summer school. For more mathematically inclined users, there are regular Winter Schools in Nice and in 2017 there was a school on the use of the Univalent Foundations library in Birmingham.

Since 2016, Coq also provides a central repository for Coq packages, the Coq opam archive, relying on the OCaml opam package manager and including around 250 packages contributed by users. It would be too long to make a detailed list of the uses of Coq in the wild. We only highlight four research projects relying heavily on Coq. The Mathematical Components library has its origins in the formal proof of the Four Colour Theorem and has grown to cover many areas of mathematics in Coq using the now integrated (since Coq 8.7) SSReflect proof language. The DeepSpec project is an NSF Expedition project led by A. Appel whose aim is full-stack verification of a software system, from machine-checked proofs of circuits to an operating system to a web-browser, entirely written in Coq and integrating many large projects into one. The ERC CoqHoTT project led by N. Tabareau aims to use logical tools to extend the expressive power of Coq, dealing with the univalence axiom and effects. The ERC RustBelt project led by D. Dreyer concerns the development of rigorous formal foundations for the Rust programming language, using the Iris Higher-Order Concurrent Separation Logic Framework in Coq.

We next briefly describe the main components of Coq.

### *3.2.1. The underlying logic and the verification kernel*

The architecture adopts the so-called de Bruijn principle: the well-delimited *kernel* of Coq ensures the correctness of the proofs validated by the system. The kernel is rather stable with modifications tied to the evolution of the underlying Calculus of Inductive Constructions formalism. The kernel includes an interpreter of the programs expressible in the CIC and this interpreter exists in two flavours: a customisable lazy evaluation machine written in OCaml and a call-by-value bytecode interpreter written in C dedicated to efficient computations. The kernel also provides a module system.

### *3.2.2. Programming and specification languages*

The concrete user language of Coq, called *Gallina*, is a high-level language built on top of the CIC. It includes a type inference algorithm, definitions by complex pattern-matching, implicit arguments, mathematical notations and various other high-level language features. This high-level language serves both for the development of programs and for the formalisation of mathematical theories. Coq also provides a large set of commands. Gallina and the commands together forms the *Vernacular* language of Coq.

### *3.2.3. Standard library*

The standard library is written in the vernacular language of Coq. There are libraries for various arithmetical structures and various implementations of numbers (Peano numbers, implementation of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ with binary digits, implementation of $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ using machine words, axiomatisation of $\mathbb{R}$). There are libraries for lists, list of a specified length, sorts, and for various implementations of finite maps and finite sets. There are libraries on relations, sets, orders.

### *3.2.4. Tactics*

The tactics are the methods available to conduct proofs. This includes the basic inference rules of the CIC, various advanced higher level inference rules and all the automation tactics. Regarding automation, there are tactics for solving systems of equations, for simplifying ring or field expressions, for arbitrary proof search, for semi-decidability of first-order logic and so on. There is also a powerful and popular untyped scripting language for combining tactics into more complex tactics.

Note that all tactics of Coq produce proof certificates that are checked by the kernel of Coq. As a consequence, possible bugs in proof methods do not hinder the confidence in the correctness of the Coq checker. Note also that the CIC being a programming language, tactics can have their core written (and certified) in the own language of Coq if needed.

### *3.2.5. Extraction*

Extraction is a component of Coq that maps programs (or even computational proofs) of the CIC to functional programs (in OCaml, Scheme or Haskell). Especially, a program certified by Coq can further be extracted to a program of a full-fledged programming language then benefiting of the efficient compilation, linking tools, profiling tools, ... of the target language.

## 3.3. Dependently typed programming languages

Dependently typed programming (shortly DTP) is an emerging concept referring to the diffuse and broadening tendency to develop programming languages with type systems able to express program properties finer than the usual information of simply belonging to specific data-types. The type systems of dependently-typed programming languages allow to express properties *dependent* of the input and the output of the program (for instance that a sorting program returns a list of same size as its argument). Typical examples of such languages were the Cayenne language, developed in the late 90's at Chalmers University in Sweden and the DML language developed at Boston. Since then, various new tools have been proposed, either as typed programming languages whose types embed equalities (Ωmega at Portland, ATS at Boston, ...) or as hybrid logic/programming frameworks (Agda at Chalmers University, Twelf at Carnegie, Delphin at Yale, OpTT at U. Iowa, Epigram at Nottingham, ...).

DTP contributes to a general movement leading to the fusion between logic and programming. Coq, whose language is both a logic and a programming language which moreover can be extracted to pure ML code plays a role in this movement and some frameworks combining logic and programming have been proposed on top of Coq (Concoqtion at Rice and Colorado, Ynot at Harvard, Why in the ProVal team at Inria, Iris at MPI-Saarbrucken). It also connects to Hoare logic, providing frameworks where pre- and post-conditions of programs are tied with the programs.

DTP approached from the programming language side generally benefits of a full-fledged language (e.g. supporting effects) with efficient compilation. DTP approached from the logic side generally benefits of an expressive specification logic and of proof methods so as to certify the specifications. The weakness of the approach from logic however is generally the weak support for effects or partial functions.

### 3.3.1. Type-checking and proof automation

In between the decidable type systems of conventional data-types based programming languages and the full expressiveness of logically undecidable formulae, an active field of research explores a spectrum of decidable or semi-decidable type systems for possible use in dependently typed programming languages. At the beginning of the spectrum, this includes, for instance, the system F's extension $\mathrm{ML}_F$ of the ML type system or the generalisation of abstract data types with type constraints (G.A.D.T.) such as found in the Haskell programming language. At the other side of the spectrum, one finds arbitrary complex type specification languages (e.g. that a sorting function returns a list of type "sorted list") for which more or less powerful proof automation tools exist – generally first-order ones.

## 3.4. Around and beyond the Curry-Howard correspondence

For two decades, the Curry-Howard correspondence has been limited to the intuitionistic case but since 1990, an important stimulus spurred on the community following Griffin's discovery that this correspondence was extensible to classical logic. The community then started to investigate unexplored potential connections between computer science and logic. One of these fields is the computational understanding of Gentzen's sequent calculus while another one is the computational content of the axiom of choice.

### 3.4.1. Control operators and classical logic

Indeed, a significant extension of the Curry-Howard correspondence has been obtained at the beginning of the 90's thanks to the seminal observation by Griffin [85] that some operators known as control operators were typable by the principle of double negation elimination ($\neg\neg A \Rightarrow A$), a principle that enables classical reasoning.

Control operators are used to jump from one location of a program to another. They were first considered in the 60's by Landin [102] and Reynolds [108] and started to be studied in an abstract way in the 80's by Felleisen *et al* [81], leading to Parigot's $\lambda\mu$-calculus [106], a reference calculus that is in close Curry-Howard correspondence with classical natural deduction. In this respect, control operators are fundamental pieces to establish a full connection between proofs and programs.

### 3.4.2. Sequent calculus

The Curry-Howard interpretation of sequent calculus started to be investigated at the beginning of the 90's. The main technicality of sequent calculus is the presence of *left introduction* inference rules, for which two kinds of interpretations are applicable. The first approach interprets left introduction rules as construction rules for a language of patterns but it does not really address the problem of the interpretation of the implication connective. The second approach, started in 1994, interprets left introduction rules as evaluation context formation rules. This line of work led in 2000 to the design by Hugo Herbelin and Pierre-Louis Curien of a symmetric calculus exhibiting deep dualities between the notion of programs and evaluation contexts and between the standard notions of call-by-name and call-by-value evaluation semantics.

### 3.4.3. Abstract machines

Abstract machines came as an intermediate evaluation device, between high-level programming languages and the computer microprocessor. The typical reference for call-by-value evaluation of $\lambda$-calculus is Landin's SECD machine [101] and Krivine's abstract machine for call-by-name evaluation [98], [97]. A typical abstract machine manipulates a state that consists of a program in some environment of bindings and some evaluation context traditionally encoded into a "stack".

### *3.4.4. Delimited control*

Delimited control extends the expressiveness of control operators with effects: the fundamental result here is a completeness result by Filinski [82]: any side-effect expressible in monadic style (and this covers references, exceptions, states, dynamic bindings, ...) can be simulated in $\lambda$-calculus equipped with delimited control.

## 3.5. Effective higher-dimensional algebra

### *3.5.1. Higher-dimensional algebra*

Like ordinary categories, higher-dimensional categorical structures originate in algebraic topology. Indeed, $\infty$-groupoids have been initially considered as a unified point of view for all the information contained in the homotopy groups of a topological space $X$: the *fundamental* $\infty$*-groupoid*$\Pi(X)$ of $X$ contains the elements of $X$ as 0-dimensional cells, continuous paths in $X$ as 1-cells, homotopies between continuous paths as 2-cells, and so on. This point of view translates a topological problem (to determine if two given spaces $X$ and $Y$ are homotopically equivalent) into an algebraic problem (to determine if the fundamental groupoids $\Pi(X)$ and $\Pi(Y)$ are equivalent).

In the last decades, the importance of higher-dimensional categories has grown fast, mainly with the new trend of *categorification* that currently touches algebra and the surrounding fields of mathematics. Categorification is an informal process that consists in the study of higher-dimensional versions of known algebraic objects (such as higher Lie algebras in mathematical physics [65]) and/or of "weakened" versions of those objects, where equations hold only up to suitable equivalences (such as weak actions of monoids and groups in representation theory [80]).

Since a few years, the categorification process has reached logic, with the introduction of homotopy type theory. After a preliminary result that had identified categorical structures in type theory [94], it has been observed recently that the so-called "identity types" are naturally equiped with a structure of $\infty$-groupoid: the 1-cells are the proofs of equality, the 2-cells are the proofs of equality between proofs of equality, and so on. The striking resemblance with the fundamental $\infty$-groupoid of a topological space led to the conjecture that homotopy type theory could serve as a replacement of set theory as a foundational language for different fields of mathematics, and homotopical algebra in particular.

### *3.5.2. Higher-dimensional rewriting*

Higher-dimensional categories are algebraic structures that contain, in essence, computational aspects. This has been recognised by Street [112], and independently by Burroni [71], when they have introduced the concept of *computad* or *polygraph* as combinatorial descriptions of higher categories. Those are directed presentations of higher-dimensional categories, generalising word and term rewriting systems.

In the recent years, the algebraic structure of polygraph has led to a new theory of rewriting, called *higher-dimensional rewriting*, as a unifying point of view for usual rewriting paradigms, namely abstract, word and term rewriting [99], [104], [86], [87], and beyond: Petri nets [89] and formal proofs of classical and linear logic have been expressed in this framework [88]. Higher-dimensional rewriting has developed its own methods to analyse computational properties of polygraphs, using in particular algebraic tools such as derivations to prove termination, which in turn led to new tools for complexity analysis [68].

### *3.5.3. Squier theory*

The homotopical properties of higher categories, as studied in mathematics, are in fact deeply related to the computational properties of their polygraphic presentations. This connection has its roots in a tradition of using rewriting-like methods in algebra, and more specifically in the work of Anick [63] and Squier [111], [110] in the 1980s: Squier has proved that, if a monoid $M$ can be presented by a *finite*, *terminating* and *confluent* rewriting system, then its third integral homology group $H_3(M, \mathbb{Z})$ is finitely generated and the monoid $M$ has *finite derivation type* (a property of homotopical nature). This allowed him to conclude that finite convergent rewriting systems were not a universal solution to decide the word problem of finitely generated monoids. Since then, Yves Guiraud and Philippe Malbos have shown that this connection was part of a deeper unified theory when formulated in the higher-dimensional setting [14], [15], [91], [92], [93].

In particular, the computational content of Squier's proof has led to a constructive methodology to produce, from a convergent presentation, *coherent presentations* and *polygraphic resolutions* of algebraic structures, such as monoids [14] and algebras [31]. A coherent presentation of a monoid $M$ is a 3-dimensional combinatorial object that contains not only a presentation of $M$ (generators and relations), but also higher-dimensional cells, each of which corresponding to two fundamentally different proofs of the same equality: this is, in essence, the same as the proofs of equality of proofs of equality in homotopy type theory. When this process of "unfolding" proofs of equalities is pursued in every dimension, one gets a polygraphic resolution of the starting monoid $M$. This object has the following desirable qualities: it is free and homotopically equivalent to $M$ (in the canonical model structure of higher categories [100], [64]). A polygraphic resolution of an algebraic object $X$ is a faithful formalisation of $X$ on which one can perform computations, such as homotopical or homological invariants of $X$. In particular, this has led to new algorithms and proofs in representation theory [10], and in homological algebra [90][31].

# 4. Highlights of the Year

## 4.1. Highlights of the Year

### *4.1.1. Awards*

Matthieu Sozeau received a Distinguished Paper award at ICFP 2018 for his work on "Equivalences for Free!"[36], together with co-authors Nicolas Tabareau and Eric Tanter.

Amina Doumane received in January 2018 the best paper award given by *La Recherche* for her paper in LICS 2017 entitled *Constructive Completeness for the Linear-Time mu-Calculus* for which she already received the Kleene Award from the LICS conference in 2017.

Amina Doumane received the Ackermann Award from the EACSL committee. As a result, she was invited to give a lecture at CSL 2018.

# 5. New Software and Platforms

## 5.1. Coq

*The Coq Proof Assistant*
KEYWORDS: Proof - Certification - Formalisation
SCIENTIFIC DESCRIPTION: Coq is an interactive proof assistant based on the Calculus of (Co-)Inductive Constructions, extended with universe polymorphism. This type theory features inductive and co-inductive families, an impredicative sort and a hierarchy of predicative universes, making it a very expressive logic. The calculus allows to formalize both general mathematics and computer programs, ranging from theories of finite structures to abstract algebra and categories to programming language metatheory and compiler verification. Coq is organised as a (relatively small) kernel including efficient conversion tests on which are built a set of higher-level layers: a powerful proof engine and unification algorithm, various tactics/decision procedures, a transactional document model and, at the very top an IDE.
FUNCTIONAL DESCRIPTION: Coq provides both a dependently-typed functional programming language and a logical formalism, which, altogether, support the formalisation of mathematical theories and the specification and certification of properties of programs. Coq also provides a large and extensible set of automatic or semi-automatic proof methods. Coq's programs are extractible to OCaml, Haskell, Scheme, ...

RELEASE FUNCTIONAL DESCRIPTION: Coq version 8.8.2 contains the result of refinements and stabilization of features and deprecations, cleanups of the internals of the system along with a few new features.

Summary of changes:

Kernel: fix a subject reduction failure due to allowing fixpoints on non-recursive values (#407), by Matthieu Sozeau. Handling of evars in the VM (#935) by Pierre-Marie Pédrot.

Notations: many improvements on recursive notations and support for destructuring patterns in the syntax of notations by Hugo Herbelin.

Proof language: tacticals for profiling, timing and checking success or failure of tactics by Jason Gross. The focusing bracket { supports single-numbered goal selectors, e.g. 2:{, (#6551) by Théo Zimmermann.

Vernacular: cleanup of definition commands (#6653) by Vincent Laporte and more uniform handling of the Local flag (#1049), by Maxime Dénès. Experimental Show Extraction command (#6926) by Pierre Letouzey. Coercion now accepts Prop or Type as a source (#6480) by Arthur Charguéraud. Export modifier for options allowing to export the option to modules that Import and not only Require a module (#6923), by Pierre-Marie Pédrot.

Universes: many user-level and API level enhancements: qualified naming and printing, variance annotations for cumulative inductive types, more general constraints and enhancements of the minimization heuristics, interaction with modules by Gaëtan Gilbert, Pierre-Marie Pédrot and Matthieu Sozeau.

Library: Decimal Numbers library (#6599) by Pierre Letouzey and various small improvements.

Documentation: a large community effort resulted in the migration of the reference manual to the Sphinx documentation tool. The new documentation infrastructure (based on Sphinx) is by Clément Pit-Claudel. The migration was coordinated by Maxime Dénès and Paul Steckler, with some help of Théo Zimmermann during the final integration phase. The 14 people who ported the manual are Calvin Beck, Heiko Becker, Yves Bertot, Maxime Dénès, Richard Ford, Pierre Letouzey, Assia Mahboubi, Clément Pit-Claudel, Laurence Rideau, Matthieu Sozeau, Paul Steckler, Enrico Tassi, Laurent Théry, Nikita Zyuzin.

Tools: experimental -mangle-names option to coqtop/coqc for linting proof scripts (#6582), by Jasper Hugunin. Main changes:

Critical soundness bugs were fixed between versions 8.8.0 and 8.8.2, and a PDF version of the reference manual was made available. The Windows installer also includes many more external packages that can be individually selected for installation.

On the implementation side, the dev/doc/changes.md file documents the numerous changes to the implementation and improvements of interfaces. The file provides guidelines on porting a plugin to the new version.

More information can be found in the CHANGES file. Feedback and bug reports are extremely welcome.

Distribution Installers for Windows 32 bits (i686), Windows 64 bits (x8_64) and macOS are available. They come bundled with CoqIDE. Windows binaries now include the Bignums library.

Complete sources of the files installed by the Windows installers are made available, to comply with license requirements.
NEWS OF THE YEAR: Version 8.8.0 was released in April 2018 and version 8.8.2 in September 2018. This is the third release of Coq developed on a time-based development cycle. Its development spanned 6 months from the release of Coq 8.7 and was based on a public road-map. It attracted many external contributions. Code reviews and continuous integration testing were systematically used before integration of new features, with an important focus given to compatibility and performance issues.

The main advances in this version are cleanups and fixes in the many different components of the system, ranging from low level kernel fixes to advances in the support of notations and tacticals for selecting goals. A large community effort was made to move the documentation to the Sphinx format, providing a more accessible online ressource to users.

- Participants: Abhishek Anand, C. J. Bell, Yves Bertot, Frédéric Besson, Tej Chajed, Pierre Courtieu, Maxime Denes, Julien Forest, Emilio Jesús Gallego Arias, Gaëtan Gilbert, Benjamin Grégoire, Jason Gross, Hugo Herbelin, Ralf Jung, Matej Kosik, Sam Pablo Kuper, Xavier Leroy, Pierre Letouzey, Assia Mahboubi, Cyprien Mangin, Érik Martin-Dorel, Olivier Marty, Guillaume Melquiond, Pierre-Marie Pédrot, Benjamin C. Pierce, Lars Rasmusson, Yann Régis-Gianas, Lionel Rieg, Valentin

Robert, Thomas Sibut-Pinote, Michael Soegtrop, Matthieu Sozeau, Arnaud Spiwack, Paul Steckler, George Stelle, Pierre-Yves Strub, Enrico Tassi, Hendrik Tews, Laurent Théry, Amin Timany, Vadim Zaliva and Théo Zimmermann

- Partners: CNRS - Université Paris-Sud - ENS Lyon - Université Paris-Diderot
- Contact: Matthieu Sozeau
- Publication: The Coq Proof Assistant, version 8.8.0
- URL: http://coq.inria.fr/

## 5.2. Equations

KEYWORDS: Coq - Dependent Pattern-Matching - Proof assistant - Functional programming
SCIENTIFIC DESCRIPTION: Equations is a tool designed to help with the definition of programs in the setting of dependent type theory, as implemented in the Coq proof assistant. Equations provides a syntax for defining programs by dependent pattern-matching and well-founded recursion and compiles them down to the core type theory of Coq, using the primitive eliminators for inductive types, accessibility and equality. In addition to the definitions of programs, it also automatically derives useful reasoning principles in the form of propositional equations describing the functions, and an elimination principle for calls to this function. It realizes this using a purely definitional translation of high-level definitions to core terms, without changing the core calculus in any way, or using axioms.
FUNCTIONAL DESCRIPTION: Equations is a function definition plugin for Coq (supporting Coq 8.6 and 8.7), that allows the definition of functions by dependent pattern-matching and well-founded, mutual or nested structural recursion and compiles them into core terms. It automatically derives the clauses equations, the graph of the function and its associated elimination principle.

Equations is based on a simplification engine for the dependent equalities appearing in dependent eliminations that is also usable as a separate tactic, providing an axiom-free variant of dependent destruction. The main features of Equations include:

Dependent pattern-matching in the style of Agda/Epigram, with inaccessible patterns, with and where clauses. The use of the K axiom or a proof of K is configurable.

Support for well-founded recursion using by rec annotations, and automatic derivation of the subterm relation for inductive families.

Support for mutual and nested structural recursion using with and where auxilliary definitions, allowing to factor multiple uses of the same nested fixpoint definition. It proves the expected elimination principles for mutual and nested definitions.

Automatic generation of the defining equations as rewrite rules for every definition.

Automatic generation of the unfolding lemma for well-founded definitions (requiring only functional extensionality).

Automatic derivation of the graph of the function and its elimination principle. In case the automation fails to prove these principles, the user is asked to provide a proof.

A new dependent elimination tactic based on the same splitting tree compilation scheme that can advantageously replace dependent destruction and sometimes inversion as well. The as clause of dependent elimination allows to specify exactly the patterns and naming of new variables needed for an elimination.

A set of Derive commands for automatic derivation of constructions from an inductive type: its signature, no-confusion property, well-founded subterm relation and decidable equality proof, if applicable.

NEWS OF THE YEAR: Equations 1.0 was released in december this year, after 7 years of (non-continuous) development. It provides the first feature-full version of the software. It has been tried and tested on small to medium scale examples (available on the website). Equations was presented at the Type Theory Tools EUTypes meeting in January 2017 in Paris, and another demo/presentation will be given at PEPM 2018 in Los Angeles in January 2018.

- Participants: Matthieu Sozeau and Cyprien Mangin
- Contact: Matthieu Sozeau
- Publications: Equations reloaded - Equations for Hereditary Substitution in Leivant's Predicative System F: A Case Study - Equations: A Dependent Pattern-Matching Compiler
- URL: http://mattam82.github.io/Coq-Equations/

## 5.3. Rewr

*Rewriting methods in algebra*
KEYWORDS: Computer algebra system (CAS) - Rewriting systems - Algebra
FUNCTIONAL DESCRIPTION: Rewr is a prototype of computer algebra system, using rewriting methods to compute resolutions and homotopical invariants of monoids. The library implements various classical constructions of rewriting theory (such as completion), improved by experimental features coming from Garside theory, and allows homotopical algebra computations based on Squier theory. Specific functionalities have been developed for usual classes of monoids, such as Artin monoids and plactic monoids.
NEWS OF THE YEAR: Rewr has been extended with the experimental KGB completion algorithm, based on Knuth-Bendix completion procedure improved by techniques coming from Garside theory.

- Participants: Yves Guiraud and Samuel Mimram
- Contact: Yves Guiraud
- Publications: Higher-dimensional categories with finite derivation type - Higher-dimensional normalisation strategies for acyclicity - Coherent presentations of Artin monoids - A Homotopical Completion Procedure with Applications to Coherence of Monoids - Polygraphs of finite derivation type - Quadratic normalisation in monoids
- URL: http://www.lix.polytechnique.fr/Labo/Samuel.Mimram/rewr

## 5.4. Catex

KEYWORDS: LaTeX - String diagram - Algebra
FUNCTIONAL DESCRIPTION: Catex is a Latex package and an external tool to typeset string diagrams easily from their algebraic expression. Catex works similarly to Bibtex.
NEWS OF THE YEAR: It is now possible to add labels to objects and morphisms

- Participant: Yves Guiraud
- Contact: Yves Guiraud
- URL: https://www.irif.fr/~guiraud/catex/catex.zip

## 5.5. Cox

KEYWORDS: Computer algebra system (CAS) - Rewriting systems - Algebra
FUNCTIONAL DESCRIPTION: Cox is a Python library for the computation of coherent presentations of Artin monoids, with experimental features to compute the lower dimensions of the Salvetti complex.

- Participant: Yves Guiraud
- Contact: Yves Guiraud
- Publications: Coherent presentations of Artin monoids - A Homotopical Completion Procedure with Applications to Coherence of Monoids
- URL: https://www.irif.fr/~guiraud/cox/cox.zip

# 6. New Results

## 6.1. Effects in proof theory and programming

**Participants:** Hugo Herbelin, Yann Régis-Gianas, Alexis Saurin, Exequiel Rivas Gadda.

### *6.1.1. Interfaces for computational effects*

Exequiel Rivas studied the relation between interfaces for computational effects in programming languages: arrows, idioms and monads. Building on previous results of Lindley, Yallop and Wadler, a categorical account was developed by means of monoidal adjunctions. This work was presented in MSFP 2018 [40] and later in SYCO I. Together with Ruben Pieters and Tom Schrijvers, a journal version of the article is currently being prepared that includes this work and previous work on non-monadic handlers. It will be submitted to the Journal of Functional Programming.

### *6.1.2. Monads with merging*

In collaboration with Mauro Jaskelioff, Exequiel Rivas developed monads with merge-like operators. These operators are based on two well-known algebraic theories for concurrency: classic process algebras and the more recent concurrent monoids. This resulted in an article submitted to FoSSaCS.

### *6.1.3. Relative effects: coherence for skew structures*

In joint work with Mauro Jaskelioff, Tarmo Uustalu and Niccolò Veltri, Exequiel Rivas developed coherence theorems in the setting of categories with skew structures: skew monoidal categories, skew near-rig categories, skew semigroup categories. These skew structures are motivated by the study of relative effects in programming languages, where the primary example are relative monads. The results are formalised in the programming language Agda. A journal article is currently being written.

### *6.1.4. Effectful proving*

Hugo Herbelin started a program of reconstruction of different levels of computational strength of logic by means of translation to a core logic of polarised linear connectives.

### *6.1.5. On the computational strength of choice axioms*

With the goal of transferring the effectful computational contents of the dependent choice to other forms of choice or bar induction axioms, Hugo Herbelin worked at clarifying the folklore regarding the strengths of various forms of choice and of bar induction.

In collaboration with Boban Velickovic, Alexis Saurin advised the LMFI master internship of Ikram Cherigui on classical realisability and forcing in set theory.

### *6.1.6. Effectful systems in Coq*

In collaboration with Thomas Letan (Agence Nationale pour la Sécurité des Systèmes Informatiques), Pierre Chifflier (ANSSI) and Guillaume Hiet (Centrale Supélec), Yann Régis-Gianas developed a new approach to model and verify effectful systems in Coq. This work has been presented at FM 2018 [38].

## 6.2. Reasoning and programming with infinite data

**Participants:** Yann Régis-Gianas, Alexis Saurin, Abhishek De, Luc Pellissier, Xavier Onfroy.

This theme is part of the ANR project Rapido (see the National Initiatives section) which goes until end of september 2019.

### *6.2.1. Proof theory of infinitary and circular proofs*

In collaboration with David Baelde, Amina Doumane, Guilhem Jaber and Denis Kuperberg, Alexis Saurin extended the proof theory of infinite and circular proofs for fixed-point logics in various directions by relaxing the validity condition necessary to distinguish sound proofs from invalid ones. The original validity condition considered by Baelde, Doumane and Saurin in CSL 2016 rules out lots of proofs which are computationally and semantically sound and does not account for the cut-axiom interaction in sequent proofs.

In the setting of sequent calculus, Saurin introduced together with Baelde, Doumane and Jaber a relaxed validity condition to allow infinite branches to be supported by threads bouncing on axioms and cuts. This allows for a much more flexible criterion, inspired from Girard's geometry of interaction. The most general form of this criterion does not ensure productivity due to a discrepancy between the sequential nature of proofs in sequent calculus and the parallel nature of threads. Several directions of research have therefore been investigated from that point:

- In sequent calculus, Baelde, Doumane and Saurin provided a slight restriction of the full bouncing validity which grants productivity and validity of the cut-elimination process. This restriction still strictly extends previous notions of validity and is actually expressive enough to be undecidable as proved together with Kuperberg. Decidability can be recovered by constraining the shapes of bounces. Doumane and Saurin were able in the fall 2018 to generalise the CSL proof technique to be applicable to bouncing threads. Those results are currently being written targetting a submission early 2019.
- In the setting of natural deduction, Saurin and Jaber introduced a validity criterion aiming at ensuring productivity of a circular $\lambda$-calculus with inductive and coinductive types.
- In the fall 2018, Abhishek De started his PhD under Saurin's supervision. The first part of his PhD work is dedicated to lifting the proof theory of circular and infinitary proofs to the setting of proof nets, in which the bouncing criterion will be much more convenient to work with since the discrepancy between sequent proofs and parallel threads will be dealt with.

### *6.2.2. Brotherston-Simpson's conjecture: Finitising circular proofs*

An important and most active research topic on circular proofs is the comparison of circular proof systems with usual proof systems with induction and co-induction rules à la Park. This can be viewed as comparing the proof-theoretical power of usual induction reasoning with that of Fermat's infinite descent method. Berardi and Tatsuta, as well as Simpson, obtained in 2017 important results in this direction for logics with inductive predicates à la Martin-Löf. Those frameworks, however, are weaker than those of fixpoint logic which can express and mix least and greatest fixpoints by interleaving $\mu$ and $\nu$ statements. New results on this topics followed in 2018.

In a work with Nollet and Tasson, Saurin published in CSL 2018 a new validity condition which is quite straightfoward to check (it can be checked at the level of elementary cycles of the circular proofs, while the other criteria need to check a condition on every infinite branch) and still capture all circular proofs obtained from $\mu MALL$ finite proofs [46]. The condition for cycling in those proofs is more constrained than that of Baelde, Doumane and Saurin, but the proof contains more information which can be used to exctract inductive invariants. With this validity condition which can be useful for proof search for circular proofs, they obtained partial finitisation results and are currently aiming at solving the most general Brotherston-Simpson's conjecture.

### *6.2.3. Streams and classical logic*

Luc Pellissier started a post-doc in december 2018 funded by the RAPIDO project and started working with Alexis Saurin on the stream interpretation of $\Lambda\mu$-calculi by investigating the connection between $\Lambda\mu$-calculus and the parsimonious $\lambda$-calculus.

### 6.2.4. Formalising circular proofs and their validity condition

During the spring and summer 2018, Saurin started with Xavier Onfroy a formalisation of circular proofs in Coq. Until now, Onfroy formalised parity-automata and their meta-theory as a first step to capture the decidability condition of circular proofs. Preliminary formalisations of circular proofs have been considered by Onfroy but shall still be pursued in order to fit into the picture.

## 6.3. Effective higher-dimensional algebra

**Participants:** Antoine Allioux, Pierre-Louis Curien, Eric Finster, Yves Guiraud, Cédric Ho Thanh, Matthieu Sozeau.

### 6.3.1. Rewriting methods in algebra

Yves Guiraud has written with Philippe Malbos (Univ. Lyon 1) a survey on the use of rewriting methods in algebra, centered on a formulation of Squier's homotopical and homological theorems in the modern language of higher-dimensional categories. This article is intended as an introduction to the domain, mainly for graduate students, and has appeared in Mathematical Structures in Computer Science [32].

Yves Guiraud has completed a four-year collaboration with Eric Hoffbeck (Univ. Paris 13) and Philippe Malbos (Univ. Lyon 1), whose aim was to develop a theory of rewriting in associative algebras, with a view towards applications in homological algebra. They adapted the known notion of polygraph [71] to higher-dimensional associative algebras, and used these objects to develop a rewriting theory on associative algebras that generalises the two major tools for computations in algebras: Gröbner bases [70] and Poincaré-Birkhoff-Witt bases [107]. Then, they transposed the construction of [14], based on an extension of Squier's theorem [110] in higher dimensions, to compute small polygraphic resolutions of associative algebras from convergent presentations. Finally, this construction has been related to the Koszul homological property, yielding necessary or sufficient conditions for an algebra to be Koszul. The resulting work will appear in Mathematische Zeitschrift [31].

Yves Guiraud has written his "Habilitation à diriger des recherches" manuscript, as a survey on rewriting methods in algebra based on Squier theory [13]. The defense is planned for Spring 2019.

Yves Guiraud works with Dimitri Ara (Univ. Aix-Marseille), Albert Burroni, Philippe Malbos (Univ. Lyon 1), François Métayer (Univ. Nanterre) and Samuel Mimram (École Polytechnique) on a reference book on the theory of polygraphs and higher-dimensional categories, and their applications in rewriting theory and homotopical algebra.

Yves Guiraud works with Marcelo Fiore (Univ. Cambridge) on the theoretical foundations of higher-dimensional algebra, in order to develop a common setting to develop rewriting methods for various algebraic structures at the same time. Practically, they aim at a definition of polygraphic resolutions of monoids in monoidal categories, based on the recent notion of $n$-oid in an $n$-oidal category. This theory will subsume the known cases of monoids and associative algebras, and encompass a wide range of objects, such as Lawvere theories (for term rewriting), operads (for Gröbner bases) or higher-order theories (for the $\lambda$-calculus).

Opetopes are a formalisation of higher many-to-one operations leading to one of the approaches for defining weak $\omega$-categories. Opetopes were originally defined by Baez and Dolan. A reformulation (leading to a more carefully crafted definition) has been later provided by Batanin, Joyal, Kock and Mascari, based on the notion of polynomial functor. Pierre-Louis Curien, Cédric Ho Thanh and Samuel Mimram have developped (in several variants) a type-theoretical treatment of opetopes and finite opetopic sets, and have shown that the models of their type theory are indeed the opetopic sets as defined mathematically by the above authors. This work is being submitted to an international conference. Also, Cédric Ho Thanh has given a direct precise proof of the equivalence between many-to-one polygraphs and opetopic sets, thus establishing a connection with the theory of polygraphs [57].

### *6.3.2. Garside methods in algebra and rewriting*

Building on [9], Yves Guiraud is currently finishing with Matthieu Picantin (Univ. Paris 7) a work that generalises already known constructions such as the bar resolution, several resolutions defined by Dehornoy and Lafont [79], and the main results of Gaussent, Guiraud and Malbos on coherent presentations of Artin monoids [10], to monoids with a Garside family. This allows an extension of the field of application of the rewriting methods to other geometrically interesting classes of monoids, such as the dual braid monoids.

Still with Matthieu Picantin, Yves Guiraud develops an improvement of the classical Knuth-Bendix completion procedure, called the KGB (for Knuth-Bendix-Garside) completion procedure. The original algorithm tries to compute, from an arbitrary terminating rewriting system, a finite convergent presentation, by adding relations to solve confluence issues. Unfortunately, this algorithm fails on standard examples, like most Artin monoids with their usual presentations. The KGB procedure uses the theory of Tietze transformations, together with Garside theory, to also add new generators to the presentation, trying to reach the convergent Garside presentation identified in [9]. The KGB completion procedure is partially implemented in the prototype Rewr, developed by Yves Guiraud and Samuel Mimram.

### *6.3.3. Foundations and formalisation of higher algebra*

Antoine Allioux (PhD started in February), Eric Finster, Yves Guiraud and Matthieu Sozeau are exploring the development of higher algebra in type theory. To formalise higher algebra, one needs a new source of coherent structure in type theory. Finster has developed an internalisation of polynomial monads (of which opetopes and $\infty$-categories are instances) in type theory, which ought to provide such a coherent algebraic structure, inspired by the work of Kock et al [96]. Antoine Allioux is focusing on building an equivalence of types between categories seen as polynomial monads and the standard univalent categories in Homotopy Type Theory [22]. Another result that should follow is the ability to define simplicial types in Homotopy Type Theory, a long standing open problem in the field. An article on this subject is in preparation. Once armed with such a definition mechanism for higher algebraic structures and their algebras, it should be possible to internalise results from higher rewriting theory in type theory, which was the initial goal of this project.

### *6.3.4. Type Theory and Higher Topos Theory*

Eric Finster explored the connections between intensional type theory and the theory of higher topoi, as developed in the works on Joyal and Lurie [103]. In particular, in collaboration with Mathieu Anel, André Joyal and Georg Biedermann, he gave a proof of a new result about the generation of left exact modalities in higher topoi, which has a corresponding internalisation in Homotopy Type Theory. Applications of this result to the Goodwillie Calculus, an advanced technique in abstract homotopy theory, resulted in the article [28].

## 6.4. Incrementality

**Participants:** Thibaut Girka, Yann Régis-Gianas.

### *6.4.1. Incrementality in proof languages*

In collaboration with Paolo Giarrusso, Philipp Shuster and Yufei Cai (Univ Marburg, Allemagne), Yann Régis-Gianas developed a new method to incrementalise higher-order programs using formal derivatives and static caching. Yann Régis-Gianas has developed a mechanised proof for this transformation as well as a prototype language featuring efficient derivatives for functional programs. A paper has been submitted to ESOP 2019.

In collaboration with Olivier Martinot (Paris Diderot), Yann Régis-Gianas studied a new technique to implement incrementalised operations on lists. A paper is to be submitted to ICFP 2019.

### *6.4.2. Difference languages*

Kostia Chardonnet and Yann Régis-Gianas started the formalisation of difference languages for Java, using the framework developed by Thibaut Girka. In particular, Kostia Chardonnet implemented a mechanised small step operational semantics for a large subset of Java. A paper is in preparation.

## 6.5. Metatheory and development of Coq

**Participants:** Hugo Herbelin, Pierre Letouzey, Yann Régis-Gianas, Matthieu Sozeau, Gaëtan Gilbert, Cyprien Mangin, Théo Winterhalter, Théo Zimmermann, Thierry Martinez.

### *6.5.1. Homotopy type theory*

Hugo Herbelin developed the syntax for a variant of Cohen, Coquand, Huber and Mörtberg's Cubical Type Theory where equality on types is defined to be equivalence of types, thus satisfying univalence by construction.

### *6.5.2. Proof irrelevance and Homotopy Type Theory*

Gaëtan Gilbert (PhD student of N. Tabareau, Gallinette and M. Sozeau) continued developing the theory and implementation of *strict* propositions in the calculus of inductive constructions. In collaboration with Jesper Cockx (Chalmers), they developed this notion in full in an article at POPL 19 [30]. Strict propositions enjoy definitional proof-irrelevance and are compatible with both Univalence and Uniqueness of Identity Proofs, providing a foundation for further research in both directions: dealing with strict structures in homotopy type theory, and improving the support for programming with dependent types and proofs. They have shown in particular how to translate inductive types that can be seen as strict propositions into recursively defined types, providing a fix to the "singleton elimination" criterion used in Coq to treat the interaction of propositions (in Prop) and informative objects (in Type). Together with Pierre Letouzey, Matthieu Sozeau is pursuing an adaptation of the Prop sort informed by this new result. In particular, Pierre Letouzey is now experimenting with alternative ways to handle the accessibility arguments of Coq general fixpoints during extraction. Historically, the elimination of these arguments was a consequence of the accessibility inductive type being in Prop. But this can actually be seen as a more general dead-code elimination method. This leverages the need for accessibility to be in sort Prop, and hence opens new prospects concerning the Prop universe and the proof irrelevance.

### *6.5.3. Extensionality and Intensionality in Type Theory*

Théo Winterhalter, Nicolas Tabareau and Matthieu Sozeau studied and formalised a complete translation from Extensional to Intensional Type Theory in Coq, now published at CPP 2019 [43]. They show that, contrary to the original paper proof of Oury, the target intensional type theory only needs to be extended with the Uniqueness of Identity Proofs principle and Functional Extensionality, settling concretely and formally a question that was studied semantically and up-to now only on paper by Hofmann and Altenkirch [61]. The translation was formalised using the Template-Coq framework and gives rise to an executable translation from partial terms of ETT into terms of Coq annotated with transports of equalities. This provides a simple way to justify the consistency of type theories extending the definitional equality relation by provable propositional equalities, and shows the equivalence of 2-level type theory [62] and the Homotopy Type System proposed by Voevodsky.

### *6.5.4. Dependent pattern-matching and recursion*

Cyprien Mangin and Matthieu Sozeau have continued work on the Equations plugin of Coq, Equations now provides means to define nested, mutual and well-founded recursive definitions, together with a definitional compilation of dependent-pattern matching avoiding the use of axioms. In recent work, Matthieu Sozeau uncovered a new way to deal with dependent pattern-matching on inductive families avoiding more uses of the K axiom, inspired by the work of Cockx [74], that integrates well with the simplification engine developed for Equations. An article describing this work is in revision [58].

Thierry Martinez continued the implementation of a dependent pattern-matching compilation algorithm in Coq based on the PhD thesis work of Pierre Boutillier and on the internship work of Meven Bertrand. The algorithm based on small inversion and generalisation is the object of a paper to be submitted to the TYPES post-proceedings.

### *6.5.5. Explicit Cumulativity*

Pierre Letouzey continued exploring with the help of Matthieu Sozeau a version of Coq's logic (CIC) where the cumulativity rule is explicit. This cumulativity rule is a form of coercion between Coq universes, and is done silently in Coq up to now. Having a version of CIC where the use of the cumulativity bewteen Prop and Type is traceable would be of great interest. In particular this would lead to a solid ground for the Coq extraction tool and solve some of its current limitations. Moreover, an explicit cumulativity would also help significantly the studies of Coq theoretical models. A prototype version of Coq is now available, but only a fragment of the standard library has been adapted to explicit cumulativity. In particular, the equalities of equalities currently need some amending, and this process is quite cumbersome.

### *6.5.6. Cumulativity for Inductive Types*

Together with Amin Timany, Matthieu Sozeau developed the Calculus of Cumulative Inductive Constructions which extends the cumulativity relation of universes to universe polymorphic inductive types. This work was presented at FSCD 2018 [42]. The development of the model of this calculus suggested a refinement of the implementation which was integrated in Coq 8.8, providing a more flexible subtyping relation on inductive types in Coq. Notably, this work shrinks the gap to emulate the so-called "template" polymorphism of Coq with cumulative universe polymorphism. Cumulative Inductive Types also provide an apropriate basis to formalise the notions of small and large categories in type theory, avoiding the introduction of coercions. In particular, it provides a way to define a well-behaved category of types and functions and constructions on it, like the Yoneda embedding, which would not be expressible without cumulativity. Finally, Cumulative Inductive Types allow the definition of syntactic models of type theories with cumulativity inside Coq, as pioneered by Boulier *et al* [69].

### *6.5.7. Mathematical notations in Coq*

Hugo Herbelin developed new extensions of the system of mathematical notation of Coq: support for autonomous auxiliary grammars, support for binders over arbitrary patterns, support for generic notations for applications.

### *6.5.8. Software engineering aspects of the development of Coq*

Théo Zimmermann has studied software engineering and open collaboration aspects of the development of Coq.

Following the migration of the Coq bug tracker from Bugzilla to GitHub which he conducted in 2017, he analyzed data (extracted through the GitHub API), in collaboration with Annalí Casanueva Artís from the Paris School of Economics. The results show an increased number of bugs by core developers and an increased diversity of the people commenting bug reports. These results validate *a posteriori* the usefulness of such a switch. A paper [60] has been written and has been presented at the EAQSE workshop (without proceedings). The current objective is to publish the paper in the MSR 2019 conference.

Following discussions dating back from the end of 2017, he has founded the coq-community GitHub organisation in July 2018. This is a project for a collaborative, community-driven effort for the long-term maintenance and advertisement of Coq packages. Already 10 pre-existing Coq projects (plugins and libraries) have been moved to this organisation since then (seven of them are former Coq contribs that were fixed from time to time by the Coq developers themselves – mostly by Hugo Herbelin). The organisation also hosts a "manifesto" repository for general discussion, documentation and advice to developers (including already a few reusable templates for Coq projects), and a docker-coq project to provide reusable Docker images with Coq. The next objectives are to get started on the collaborative documentation (starting with a work by Pierre Castéran from LaBRI) and to create an editorial committee. Théo Zimmermann and Yann Régis-Gianas are preparing an article of the model proposed by the various existing *-community GitHub organisations (including the elm-community organisation from which coq-community was inspired, and ocaml-community which was influenced by coq-community itself).

In addition, Théo Zimmermann has coordinated efforts to improve the documentation of Coq, has documented the release process that he had put in place with Maxime Dénès, and has developed a GitHub / GitLab bot (in OCaml) that is used to automatise many useful functions for the Coq development (continuous integration and backporting of pull requests in particular). The goal is to make this bot modular and reusable for other projects.

### 6.5.9. Coordination of the development of Coq

The amount of contributions to the Coq system increased significantly in the recent years (around 50 pull-requests are reviewed, discussed and merged each month, approximately). Hugo Herbelin, Matthieu Sozeau and Théo Zimmermann, helped by members from Gallinette (Nantes) and Marelle (Sophia-Antipolis), devoted an important part of their time to coordinate the development, to review propositions of extensions of Coq from external and/or young contributors, and to propose themselves extensions (see the corresponding paragraphs).

## 6.6. Formalisation and verification

**Participants:** Pierre-Louis Curien, Kailiang Ji, Pierre Letouzey, Jean-Jacques Lévy, Cyprien Mangin, Daniel de Rauglaudre, Matthieu Sozeau.

### 6.6.1. Proofs and surfaces

Following ideas of J. Richter-Gebert, Pierre-Louis Curien, together with Jovana Obradović (former PhD student of the team and now postdoc in Prague), joined a project with Zoran Petrić and other Serbian colleagues on formalising proofs of incidence theorems (arising by repeated use of Menelaus theorem) by means of a cyclic sequent calculus, by which is meant that a (proof of a) sequent $\vdash \Gamma$ stands for the conjunction of all (proofs of) traditional sequents $\Gamma \setminus \psi \vdash \psi$. We have designed a proof system, showed its soundness, and experimented it on an extended set of examples from elementary projective geometry. A paper is being written.

### 6.6.2. Hofstadter nested recursive functions and Coq

Pierre Letouzey continued this year the study of a family of nested recursive functions proposed by D. Hofstadter in his book “Gödel Escher Bach”. This is a generalisation of the earlier work [20], bringing a large number of new insights as well as many new conjectures. Most of the work is already certified in Coq, with generalised and/or nicer proofs, see https://www.irif.fr/~letouzey/hofstadter_g/. Many interactions with Fibonacci numbers or similar recursive sequence have been found. Pierre Letouzey even stumbled upon a Rauzy fractal during this investigation, which is still ongoing.

### 6.6.3. Real Numbers in Coq

The present Coq library of real numbers is made of 17 axioms. Daniel de Rauglaudre has been studying the possibility of making an implementation with one only axiom: the Limited Principle of Omniscience (LPO) which says that we can differentiate an infinite sequence of 0s from an infinite sequence holding something else than 0 (it seems obvious but it cannot be proved in constructive logic). This axiom had been already used in the formal proof of Puiseux’ theorem done some years ago (only axiom of this proof too).

Real numbers are defined by an infinite sequence of digits and the operations of addition and multiplication by algorithms using LPO.

It was tested in OCaml, the axiom being replaced by a function having a limit corresponding to the precision of the computation and it seems to work. But the proof in Coq that this implementation is a field stumbles on difficulties about the associativity of addition which is more complicated than expected. Several tracks have been experimented with Hugo Herbelin’s help.

### *6.6.4. Proofs of algorithms on graphs*

Jean-Jacques Lévy and Chen Ran (a PhD student at the Institute of Software, Beijing) pursue their work about formal proofs of graph algorithms. Their goal is to provide proofs of algorithms checked by computer and human readable. If these kinds of proofs exist for algorithms on inductive structures or recursive algorithms on arrays, they seem less easy to design for combinatorial structures such as graphs. In 2016, they completed proofs for algorithms computing the strongly connected components in graphs (Kosaraju - 1978 and Tarjan - 1972). Their proofs use the multi-sorted first-order logic with inductive predicates of the Why3 system (research-team Toccata, Saclay). They also widely use the numerous automatic provers interfaced with Why3. A very minor part of these proofs is also achieved in Coq. The difficulty of this approach is to combine automatic provers and the intuitive design. Another point is to define the good level of abstraction in order to avoid too many implementation features while keeping an effective presentation.

In 2017, the same proofs were fully completed in Coq-ssreflect with the Mathematical Components library by Cohen and Théry (research-team Marelle, Sophia-Antipolis), and in Isabelle-HOL by Merz (research-team VeriDis, Nancy), both proofs with the assistance of J.-J. Lévy. These proofs are between a factor 3 to 8 in length with respect to the initial Why3 proofs, but more importantly they look less human readable, mainly because of the absence of automatic deduction and several technicalities about termination. On the way, this collaboration led to a new, better presentation of the Why3 proof.

Part of this work (Tarjan 1972) was presented at JFLA 2017, a more comprehensive version was presented at the VSTTE 2017 conference in Heidelberg. Scripts of proofs can be found at http://jeanjacqueslevy.net/why3, where other proofs of graph algorithms are also present: acyclicity test, articulation points, biconnected components. A proof of Tarjan's planarity test is also under design. A paper entitled "Formal Proofs of Tarjan's Algorithm in Why3, Coq and Isabelle" is under submission to a conference.

### *6.6.5. Certified compilation and meta-programming*

Matthieu Sozeau participates to the CertiCoq project (https://www.cs.princeton.edu/~appel/certicoq) whose aim is to verify a compiler from Coq's Gallina language down to CompCert C-light which provides itself a certified compilation path to assembly language. Matthieu Sozeau focused on the front-end part of CertiCoq, providing formal proofs of the first two phases of the compiler. The first phase translates from Coq syntax to a more amenable representation for metatheoretical study, and the second phase performs extraction to an untyped lambda-calculus with datatypes and mutual (co-)fixpoints. These two phases are of general use and are now integrated and developed in the MetaCoq project. The CertiCoq team expects to release a first version of the compiler in the beginning of 2019, along with an article describing it.

MetaCoq is a project led by Matthieu Sozeau, in collaboration with Simon Boulier and Nicolas Tabareau in Nantes, Abhishek Anand and Gregory Malecha (BedRock Systems, Inc) and Yannick Forster in Saarbrucken. The project was born from the extension of the Template-Coq reification plugin of G. Malecha, which now contains:

- A specification of the typing rules of Coq and its basic metatheoretical properties (weakening, substitution). This specification is not entirely complete yet, as the positivity and guard-checking of definitions is missing. Cyprien Mangin has formalised the regular tree structure used by the guard checker, and a simple positivity check for inductive types. Its integration is ongoing.
- A (partial) proof of the correctness and completeness of a reference type-checker with respect to these rules.
- An implementation of the extraction phase of Coq, which is used in the CertiCoq project. The proof of "syntactic" correctness of this phase, that is the preservation of weak call-by-value reduction by extraction is ongoing.
- A monad giving the ability to program arbitrary plugins in Coq itself, in the style of MTac.

. The foundation of this project was published at ITP 2018 [37], and a journal article is in preparation.

In collaboration with Jan-Oliver Kaiser (MPI-SWS), Beta Ziliani (CONICET/FAMAF), Robbert Krebbers (ICIS) and Derek Dreyer (MPI-SWS), Yann Régis-Gianas participates in the Mtac2 project, a metaprogramming language for Coq. The new version of this language has been presented at ICFP 2018 [34]. It includes in particular in a depedently-typed variant of the LCF tactic typing discipline.

In collaboration with Xavier Denis (Paris Diderot), Yann Régis-Gianas is implementing a compiler for Mtac2.

### 6.6.6. *Equivalences for free!*

Nicolas Tabareau (Inria Nantes), Eric Tanter (U. Chile in Santiago) and Matthieu Sozeau developed a new parametricity translation for justifying the transport of programs and proofs by equivalences in type theory [36]. Inspired by the Univalence axiom, they show that every construction of type theory (minus inductive families indexed by universes) respect type equivalence, and provide a modified parametricity translation that can be used to construct the proof of invariance by equivalence of any term. This translation is engineered so that transports do not appear during this inference, allowing an easy implementation of a transfer metaprogram in type theory using type class inference. Using this metaprogram, one can automatically transport libraries of implementations and their proofs from one type to an equivalent one, including cases where dependent types are used. While the translation ultimately relies on the univalence axiom to treat universes, its use can be avoided in many cases, providing an effective translation that can be evaluated inside type theory.

### 6.6.7. *Detecting K-Synchronisability Violations*

Ahmed Bouajjani, Constantin Enea, Kailiang Ji and Shaz Qadeer introduced a bounded analysis that explores a special type of computations, called $k$-synchronous, for analyzing message passing programs. They gave a procedure for deciding *$k$-synchronisability* of a program, i.e., whether every computation is equivalent (has the same happens-before relation) to one of its $k$-synchronous computations. They also showed that reachability over $k$-synchronous computations and checking $k$-synchronisability are both PSPACE-complete. Furthermore, they introduced a class of programs called *flow-bounded* for which the problem of deciding whether there exists a $k > 0$ for which the program is $k$-synchronisable, is decidable. The *$k$-synchronisability* violation detection algorithm was implemented in Spin model checker. This work was published at CAV 2018 [48].

# 7. Partnerships and Cooperations

## 7.1. National Initiatives

Pierre-Louis Curien, Yves Guiraud, Hugo Herbelin, and Alexis Saurin are members of the GDR Informatique Mathématique, in the LHC (Logique, Homotopie, Catégories) and Scalp (Structures formelles pour le calcul et les preuves) working groups. Alexis Saurin is coordinator of the Scalp working group.

Pierre-Louis Curien, Yves Guiraud (local coordinator) and Matthieu Sozeau are members of the GDR Topologie Algébrique, federating French researchers working on classical topics of algebraic topology and homological algebra, such as homotopy theory, group homology, K-theory, deformation theory, and on more recent interactions of topology with other themes, such as higher categories and theoretical computer science.

Yves Guiraud is member of the GDR Tresses, federating French researchers working on algebraic, algorithmic and topological aspects of braid groups, low-dimensional topology, and connected subjects.

Yann Régis-Gianas collaborates with Mitsubishi Rennes on the topic of differential semantics. This collaboration led to the CIFRE grant for the PhD of Thibaut Girka.

Yann Régis-Gianas collaborates with ANSSI on the topic of certified ful programming in Coq.

Yann Régis-Gianas is a member of the ANR COLIS dedicated to the verification of Linux Distribution installation scripts. This project is joint with members of VALS (Univ Paris Sud) and LIFL (Univ Lille).

Yann Régis-Gianas and Alexis Saurin (coordinator) are members of the four-year RAPIDO ANR project, started in January 2015. RAPIDO aims at investigating the use of proof-theoretical methods to reason and program on infinite data objects. The goal of the project is to develop logical systems capturing infinite proofs (proof systems with least and greatest fixpoints as well as infinitary proof systems), to design and to study programming languages for manipulating infinite data such as streams both from a syntactical and semantical point of view. Moreover, the ambition of the project is to apply the fundamental results obtained from the proof-theoretical investigations (i) to the development of software tools dedicated to the reasoning about programs computing on infinite data, *e.g.* stream programs (more generally coinductive programs), and (ii) to the study of properties of automata on infinite words and trees from a proof-theoretical perspective with an eye towards model-checking problems. Other permanent members of the project are Christine Tasson from IRIF (PPS team), David Baelde from LSV, ENS-Cachan, and Pierre Clairambault, Damien Pous and Colin Riba from LIP, ENS-Lyon.

Matthieu Sozeau is a member of the CoqHoTT project led by Nicolas Tabareau (Gallinette team, Inria Nantes & École des Mines de Nantes), funded by an ERC Starting Grant. The post-doctoral grant of Eric Finster is funded by the CoqHoTT ERC and Amin Timany's 2-month visit was funded on the ERC as well.

## 7.2. European Initiatives

### *7.2.1. Collaborations in European Programs, Except FP7 & H2020*

Hugo Herbelin is a deputy representative of France in the COST action EUTYPES. The full name of the project (whose scientific leader is Herman Geuvers, from the University of Nijmegen) is "European research network on types for programming and verification".

Presentation of EUTYPES: Types are pervasive in programming and information technology. A type defines a formal interface between software components, allowing the automatic verification of their connections, and greatly enhancing the robustness and reliability of computations and communications. In rich dependent type theories, the full functional specification of a program can be expressed as a type. Type systems have rapidly evolved over the past years, becoming more sophisticated, capturing new aspects of the behaviour of programs and the dynamics of their execution. This COST Action will give a strong impetus to research on type theory and its many applications in computer science, by promoting (1) the synergy between theoretical computer scientists, logicians and mathematicians to develop new foundations for type theory, for example as based on the recent development of "homotopy type theory", (2) the joint development of type theoretic tools as proof assistants and integrated programming environments, (3) the study of dependent types for programming and its deployment in software development, (4) the study of dependent types for verification and its deployment in software analysis and verification. The action will also tie together these different areas and promote cross-fertilisation.

## 7.3. International Initiatives

### *7.3.1. IIL projects*

Matthieu Sozeau is part of an international collaboration network CSEC "Certified Software Engineering in Coq" funded by Inria Chile, Conicyt and the CoqHoTT ERC, which officially started in early 2018. The participants include Eric Tanter (primary investigator) and Nicolas Tabareau.

### *7.3.2. Inria Associate Teams Not Involved in an Inria International Labs*

#### *7.3.2.1. Associate team*

Pierre-Louis Curien and Claudia Faggian are members of the CRECOGI associate team, coordinated on one side by Ugo dal Lago (research-team FoCUS , Inria Sophia and Bologna), and on the other side by Ichiro Hasuo (NII, Tokyo). The full name of the project is Concurrent, Resourceful and full Computation, by Geometry of Interaction.

Presentation of CRECOGI: Game semantics and geometry of interaction (GoI) are two closely related frameworks whose strengh is to have the characters of both a denotational and an operational semantics. They offer a high-level, mathematical (denotational) interpretation, but are interactive in nature. The formalisation in terms of movements of tokens through which programs communicate with each other can actually be seen as a low-level program. The current limit of GoI is that the vast majority of the literature and of the software tools designed around it have a pure, sequential functional language as their source language. This project aims at investigating the application of GoI to concurrent, resourceful, and effectful computation, thus paving the way to the deployment of GoI-based correct-by-construction compilers in real-world software developments in fields like (massively parallel) high-performance computing, embedded and cyberphysical systems, and big data. The presence of both the Japanese GoI community (whose skills are centered around effects and coalgebras) and the French GoI community (more focused on linear logic and complexity analysis) bring essential, complementary, ingredients.

*7.3.2.2. Joint Inria-CAS project*

Pierre-Louis Curien is principal investigator on the French side for a joint project Inria - Chinese Academy of Sciences. The project's title is "Verification, Interaction, and Proofs". The principal investigator on the Chinese side is Ying Jiang, from the Institute of Software (ISCAS) in Beijing. The participants of the project on the French side are Pierre-Louis Curien and Jean-Jacques Lévy, as well as other members of IRIF (Thomas Ehrhard, Jean Krivine, Giovanni Bernardi, Ahmed Bouajjani, Mihaela Sighireanu, Constantin Enea, Gustavo Petri), and Gilles Dowek (Deducteam team of Inria Saclay). On the Chinese side, the participants are Ying Jiang, as well as other members of the ISCAS (Angsheng Li, Xinxin Liu, Yi Lü, Peng Wu, Yan Rongjie, Zhilin Wu, and Wenhui Zhang), and Yuxi Fu (from Shanghai Jiaotong University). The project funds the postdoc of Kailiang Ji at University Paris 7, that started in December 2017 and will end in Marche 2019.

Presentation of VIP: The line between "verification" and "proofs" is comparable to the one separating satisfiability and provability: in a formal system, a formula can be trusted either if it is satisfied in the intended model (for all of its instances), or if it can be proved formally by using the axioms and inference rules of some logical system. These two directions of work are called model-checking and proof-checking, respectively. One of the aims of the present project is to bring specialists of the two domains together and to tackle problems where model-checking and proof-checking can be combined (the "V" and the "P" of the acronym). Applications in the realm of distributed computation, or concurrency theory (the "I" of the acronym) are particularly targeted.

### *7.3.3. Inria International Partners*

*7.3.3.1. Informal International Partners*

The project-team has collaborations with University of Aarhus (Denmark), KU Leuven, University of Oregon, University of Tokyo, University of Sovi Sad and the Institute of Mathematics of the Serbian Academy of Sciences, University of Nottingham, Institute of Advanced Study, MIT, University of Cambridge, Universidad Nacional de Córdoba, and Universidad de Chile.

## 7.4. International Research Visitors

### *7.4.1. Visits of International Scientists*

Mauro Jaskelioff (National University of Rosario and CONICET, Argentina) visited the team for a week in May 2018.

Vadim Zaliva (PhD student at CMU) visited the team for one month in July 2018 and collaborated with Matthieu Sozeau on the use of Template-Coq to verify translations from shallow to deep embeddings.

### *7.4.2. Internships*

Yann Régis-Gianas supervised the internship of Loïc Peyrot (Master 1, Paris Diderot) about the development of a tool to define exercises for the learn-ocaml platform in a single ML file.

Yann Régis-Gianas supervised the internship of Carine Morel (Master 1, Paris Diderot) about the development of a user-friendly teaching-oriented documentation for the learn-ocaml platform.

Yann Régis-Gianas supervised the internship of Olivier Martinot (Licence 3, Paris Diderot) about the implementation of a set of efficient incrementalised combinators for list processing in cache-transfer style.

Alexis Saurin co-supervised the internship of Ikram Cherigi (Master 2 LMFI, Paris Diderot) about classical realisability and forcing in set theory.

Alexis Saurin supervised the internship of Xavier Onfroy (Master 2 LMFI, Paris Diderot) on formalisation of circular proofs in fixed-point logics and the decidability of validity.

Alexis Saurin supervised the internship of Kostia Chardonnet (Master 1 MPRI, Paris Diderot) about call-by-need calculus, degrees of lazyness and probabilistic lambda calculus.

### *7.4.3. Research Stays Abroad*

Pierre-Louis Curien visited East China Normal University for a month from mid-October to mid-November 2018 (collaborations with Yuxin Deng and Min Zhang) as invited professor.

Pierre-Louis Curien visited the Institute of Mathematics of the Serbian Academy of Sciences in Belgrade in September 2018 for a week (collaboration with Zoran Petrić and other coauthors).

Hugo Herbelin participated to the Types, Sets and Constructions Trimester Program at the Hausdorff Research Institute of Mathematics in Bonn, May-August 2018.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. General Chair, Scientific Chair*

Pierre-Louis Curien organised a Day of Hommage to the memory of Maurice Nivat on February 6, 2018, at University Paris 7.

Alexis Saurin organised and co-chaired with David Baelde the Paris workshop in Oxford, UK, July 7-8th 2018, collocated with FLoC 2018.

Matthieu Sozeau co-organised and co-chaired with Nicolas Tabareau the Coq Workshop 2018 in Oxford, UK, July 8th 2018, collocated with FLoC 2018.

*8.1.1.2. Member of the Organising Committees*

Yves Guiraud organised with Philippe Malbos (Univ. Lyon 1) and Samuel Mimram (École Polytechnique) the fourth edition of the workshop HDRA (Higher-Dimensional Rewriting and Algebra) in July 2018 in Oxford.

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Member of the Conference Program Committees*

Hugo Herbelin was a member of the program committee of the conference POPL 2019.

Yann Régis-Gianas was a member of the program committee of the conference PPDP 2018.

Yann Régis-Gianas was a member of the program committee of the conference JFLA 2019.

Matthieu Sozeau was member of the program committee of the conference Interactive Theorem Proving 2018 which took place in Oxford during FLoC 2018 and the 13th Workshop on Logical and Semantic Frameworks with Applications, which took place in Fortaleza, Brazil, September 26-28, 2018.

*8.1.2.2. Member of the Conference Steering Committees*

Pierre-Louis Curien is member of the steering committee of the international workshop Games for Logic and Programming Languages (GaLop).

Hugo Herbelin is a member of the steering committee of the conference TYPES.

Matthieu Sozeau is member of the steering committee of the Dependently Typed Programming international workshop (DTP).

### *8.1.3. Journal*

*8.1.3.1. Member of the Editorial Boards*

Pierre-Louis Curien is editor in chief of the Cambridge University Press journal Mathematical Structures in Computer Science (since January 2016).

Alexis Saurin is editing a special issue of MSCS dedicated to contributions in honour of Dale Miller for his 60th birthday.

*8.1.3.2. Reviewer - Reviewing Activities*

The members of the team reviewed papers for numerous journals and international conferences.

### *8.1.4. Invited Talks*

Pierre-Louis Curien gave talks on the legacy of Maurice Nivat at two special events organised to honour his memory: special sessions in the Journées du GDR IM (Ecole Polytechnique, May 2018), and at ICALP 2018 (Prague, July 2018).

Eric Finster gave an invited talk at the annual meeting of the GDR Topologie in Montpellier in October entitled "The Cotopological Tower".

Hugo Herbelin gave an invited talk on computing with Gôdel's completeness theorem using side effects at the workshop Proof, Computation and Complexity in Bonn, July 2018.

Yann Régis-Gianas gave an invited talk about copatterns in OCaml at the "Logique, Types et Preuves" workshop of the GDR GPL.

Matthieu Sozeau gave an invited talk on "The Predicative, Polymorphic, Cumulative Calculus of Inductive Constructions" at the TYPES 2018 International Conference on Types for Proofs and Programs in Braga, Portugal, 18-21 June 2018.

Matthieu Sozeau gave an invited seminar entitled "Programmer en Coq" at the Collège de France, on December 12th 2018, part of Xavier Leroy's lectures on the Curry-Howard Isomorphism.

Théo Zimmermann was invited to give a talk in the First international workshop on Empirical Answers to Questions of Software Engineering to present his work on the impact of switching bug trackers [60].

### *8.1.5. Scientific Expertise*

Pierre-Louis Curien has been an expert for a hiring committee for an assistant professor position in Logic, Computation and Programming at Stockholm University (June 2018).

Hugo Herbelin has been a reviewer for FWF (Austrian research funding agency) and NKFI (Hungarian research funding agency).

### *8.1.6. Research Administration*

Pierre-Louis Curien is a member of the Scientific Council of the CIRM (Centre International de Rencontres Mathématiques).

Pierre-Louis Curien and Yves Guiraud are members of the scientific council of the Computer Science department of University Paris 7.

Yves Guiraud is the head of the “Preuves, Programmes et Systèmes” (PPS) pole of the IRIF laboratory (since April 2016), and a member of the IRIF direction council (since September 2017).

Yann Régis-Gianas is a member of the Executive Comitee of the OCaml Foundation, acting as a representative of the teaching community.

In collaboration with Emmanuel Chailloux (UPMC), Yann Régis-Gianas is organising the next four years of IRILL, an initiative about innovation in free software.

### *8.1.7. Presentation of papers*

Pierre-Louis Curien gave a talk at the Conference “Topology in Australia and South Korea 2018”, Pohang (https://cgp.ibs.re.kr/conferences/Topology_in_Australia_and_South_Korea) in April 2018 (‘A syntactic approach to opetopes’).

Yann Régis-Gianas gave talks to present "Morbig", a static parser for POSIX Shellat at FOSDEM 2018 in Brussels, at MiniDebConf 2018 and at SLE 2018 in Boston.

Yann Régis-Gianas gave a talk at OCaml 2018 in St Louis to present Learn-OCaml, a project to support the teaching of OCaml worldwide.

Yann Régis-Gianas gave a talk at JFLA 2018 to present his work about extending OCaml with Copatterns.

Exequiel Rivas gave a talk on relating interfaces for computational effects at the Seventh Workshop on Mathematically Structured Functional Programming (MSFP 2018) in July 2018.

Exequiel Rivas gave a talk on relating interfaces for computational effects at the First Symposium on Compositional Structures (SYCO I) in September 2018.

Matthieu Sozeau gave a talk and presented a poster at PEPM 2018 on Equations, gave a talk on Typed Template Coq at CoqPL 2018, along with the traditional Coq developer session. These events were co-located with POPL 2018 in Los Angeles, CA in January 2018.

### *8.1.8. Talks in seminars*

Eric Finster gave a talk about the implementation of Catt, a proof assistant for Maltsinotis-style higher categories at the Journées pi.r2 (Fontainebleau, june 2018).

Eric Finster gave a talk on “Towards Higher Universal Algebra in Dependent Type Theory” in the working group on Higher Categories, Polygraphs and Homotopy, during the Journées PPS (November 2018).

Eric Finster gave a talk during the HoTTest Electronic Seminar on "Towards Higher Universal Algebra in Type Theory".

Eric Finster gave a talk entitled "Left Exact Modalities in Type Theory" at the Cambridge Logic and Semantics Seminar (March 2018).

Hugo Herbelin gave a talk on computing with Gödel’s completeness theorem at the seminar of the Logic team of the IMJ-PRG Paris 6 - Paris 7 lab.

Pierre Letouzey gave two talks on “Un problème d’Hofstadter pour ses lecteurs curieux” during the Journées pi.r2 (Fontainebleau, june 2018) and the Journées PPS (November 2018).

Jean-Jacques Lévy gave a talk at the IRIF Verification seminar (January 8) entitled "Proofs of graph algorithms with automation and their readability".

Jean-Jacques Lévy gave a talk at the VIP ISCAS-Inria workshop (Irif, November 19-22) entitled "Comparing a Formal Proof in Why3, Coq and Isabelle".

Yann Régis-Gianas gave a talk at Gallium seminar to present "Morbig", a static parser for POSIX Shell.

Exequiel Rivas gave a talk on “Arrows: from programming to semantics” at the Laboratoire d’Informatique de Paris Nord (LIPN), September 2018.

Exequiel Rivas gave a talk on “Interaction from monadic interfaces” during the Journées PPS, November 2018.

Exequiel Rivas gave a talk on “Interaction from monadic interfaces” at the Prosecco seminar, December 2018.

Alexis Saurin gave a talk at I2M seminar in the "logique de la programmation" group entitled "logical-by-need".

Alexis Saurin gave a talk at the VIP ISCAS-Inria workshop (Irif, November 19-22) entitled "On non-wellfounded proofs and cuts in linear logic with fixed points."

Matthieu Sozeau gave a talk on the MetaCoq Project at the VALS seminar, LRI, October 2018.

Matthieu Sozeau gave a talk on “A universe of strict propositions“ during the Journées PPS, November 2018.

### *8.1.9. Attendance to conferences, workshops, schools,...*

Hugo Herbelin attended TYPES 2018 in Porto (June), the Coq Implementors Workskop in Nice (May 2018), FLoC in Oxford (July 2018), the GPL working group in Marseille (October 2018), the Scalp working group in Saclay (November 2018).

Hugo Herbelin gave a talk on cubical type theory at the workshop Types, Homotopy Type theory and Verification (June 2018), on computing with Markov’s principle at the workshop Proof and Computation (July 2018), on a constructive proof of the axiom of dependent choice compatible with classical logic at the workshop Constructive Mathematics (August 2018), all workshops of the special trimester on types, sets and constructions in Bonn. He gave a talk on the reverse mathematics of Gödel’s completeness theorem and on the computational contents of Henkin’s proof at the seminar of the trimester.

Hugo Herbelin gave a talk on cubical type theory at the TYPES conference (June 2018).

Hugo Herbelin gave a talk on the cubical type structure of cubical type theory at the HoTT-UF workshop in Oxford, July 2018.

Jean-Jacques Lévy attended the Coq Winter School 2018-2019 (SSReflect & MathComp) at Inria Sophia-Antipolis (November 2018).

Yann Régis-Gianas attended JFLA 2018, OCaml 2018, SPLASH 2018, FOSDEM 2018 and MiniDebConf 2018.

Exequiel Rivas attended to MSFP 2018 and SYCO I.

Alexis Saurin attended FLoC 2018 in Oxford.

Matthieu Sozeau attended POPL in Los Angeles, CA (January 2018), the Coq Implementors Workshop in Nice (May 2018), the TYPES Conference in Braga, Portugal (June 2018), FLoC in Oxford (July 2018) and ICFP in St Louis, MI (September 2018).

Théo Zimmermann attented FOSDEM in Brussels (February 2018), the Coq Implementors Workskop in Nice (May 2018), FLoC in Oxford (July 2018), OpenSym in Paris (August 2018) and the EAQSE workshop in Villebrumier (November 2018).

### *8.1.10. Groupe de travail Théorie des types et réalisabilité*

This is one of the working groups of PPS, jointly organised by Hugo Herbelin and Matthieu Sozeau. The speakers in 2018 were Rodolphe Lepigre (Practical Curry-Style using Choice Operators, Local Subtyping and Circular proofs), Armaël Guéneau ( A Fistful of Dollars: Formalising Asymptotic Complexity Claims via Deductive Program Verification), Jérôme Siméon (Specifying and compiling domain specific languages using Coq: Three case studies), Laura Fontanella ’Axiom of choice in classical realisabiity), Adrien Guatto (A Generalised Modality for Recursion), Hadrien Batmalle (Preservation of properties of the original model in classical realisability), Raphaël Cauderlier (Tactics and certificates in Meta Dedukti).

### *8.1.11. Groupe de travail Catégories supérieures, polygraphes et homotopie*

Several members of the team participate actively in this weekly working group of PPS, organised by François Métayer (Univ. Nanterre) since 2009.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Master: Pierre-Louis Curien teaches in the course Models of programming languages: domains, categories, games of the MPRI (together with Thomas Ehrhard and Paul-André Melliès). Pierre-Louis Curien taught a course on the Foundations of Programming Languages at East China Normal University (12 hours, November 2018).

Master: Hugo Herbelin teaches with Nicolas Tabareau the course on Homotopy Type Theory at the LMFI.

Master: Pierre Letouzey teaches two short courses to the LMFI Master 2 students : “Programming in Coq” and “Introduction to computed-aided formal proofs”. These two courses come in addition to Pierre Letouzey’s regular duty as teacher in the Computer Science department of Paris 7 (including a course on Compilation to M2-Pro students and a course on computed-aided formal proofs to M1 students).

Master: Yann Régis-Gianas took part in the MPRI course entitled “Type systems”: he gave a 12-hour course about generalised algebraic data types, higher-order Hoare logic and dependently typed programming.

Master: Alexis Saurin taught the proof theory and lambda-calculus part of the cours fondamental de logique in M2 “Logique Mathématique et Fondements de l’Informatique”, Université Paris 7.

Alexis Saurin chairs LMFI M2 since September 2013.

Master: Matthieu Sozeau taught the MPRI course on Advanced uses of proof assistants (12 hours + a project), together with Bruno Barras (Inria Deducteam).

Matthieu Sozeau gave a guest lecture on dependent pattern-matching and Equations at the University of Saarland in April 2018.

Matthieu Sozeau gave an introductory lecture on Dependent Type Theory at the EUTYPES summer school in Ohrid, Macedonia, in August 2018.

### *8.2.2. Supervision*

Guillaume Claret defended his PhD on “Programmation avec effets en Coq” on 18 September 2018 (supervised by Hugo Herbelin and Yann Régis-Gianas). Note that the dissertation was completed in 2015 but Guillaume Claret moved in the meantime to a private company and the defense has been delayed.

PhD (completed): Thibaut Girka defended his PhD on "Differential Program Semantics" on the 3rd of July 2018, supervised by Roberto Di Cosmo and Yann Régis-Gianas.

PhD (abandoned): Cyprien Mangin, Dependent Pattern-Matching, induction-induction and higher inductive types (started in September 2015), supervised by Matthieu Sozeau and Bruno Barras. Cyprien Mangin left for industry.

PhD in progress: Théo Zimmermann (started in September 2016), supervised by Hugo Herbelin.

PhD in progress: Cédric Ho Thanh (started in September 2017), on Opetopes for higher-dimensional rewriting and koszulity, supervised by Pierre-Louis Curien and Samuel Mimram.

PhD in progress: Antoine Allioux (started in February 2018), on the formalisation of algebraic structures in type theory, supervised by Yves Guiraud and Matthieu Sozeau.

PhD in progress: Abhishek De (started in october 2018), on fixed point logics, structures for infinite proofs and their finite representations, supervised by Alexis Saurin.

The following are cosupervisions of PhD students who are not formally part of the team:

PhD in progress: Rémi Nollet, Functional reactive programming and temporal logics: their syntax and semantics - from discrete to continuous time (started in September 2016), supervised by Alexis Saurin and Christine Tasson.

PhD in progress: Gaëtan Gilbert (at Inria Nantes), Definitional proof-irrelevance in the Calculus of Inductive Constructions (started in September 2016), supervised by Nicolas Tabareau and Matthieu Sozeau.

PhD in progress: Simon Forest (at École Polytechnique), Rewriting in semistrict higher categories (started in September 2017), supervised by Yves Guiraud and Samuel Mimram.

PhD in progress: Théo Winterhalter (at Inria Nantes), Extensional to Intensional type theory and meta-theory of proof-irrelevance (started in September 2017), supervised by Nicolas Tabareau and Matthieu Sozeau.

### *8.2.3. Juries*

Pierre-Louis Curien was member of the jury of the PhD thesis of Clovis Eberhard (Université Savoie Mont Blanc), defended in June 2018.

Hugo Herbelin was a member of the jury of the PhD thesis of Andrea Vezzosi (University of Göteborg, Sweden), defended in September 2018.

Hugo Herbelin was a member of the jury of the PhD thesis of Guillaume Claret (University of Paris-Diderot), defended in September 2018.

Hugo Herbelin was referee for the PhD thesis of Simon Boulier (University of Nantes), defended in November 2018.

Hugo Herbelin was president of the jury of the PhD thesis of Pierre Cagne (University of Paris Diderot), defended in December 2018.

Yann Régis-Gianas is a member of the jury of the competitive examination for the entrance to the Ecoles Normales Supérieures and the Ecole Polytechnique.

Matthieu Sozeau was member of the jury of the PhD thesis of Amin Timany (KU Leuven, Belgium), defended in April 2018.

## 8.3. Popularisation

Pierre-Louis Curien gave a talk in the Lycée Georges Dumézil (Vernon, Eure, May 2018) on computer bugs and their prevention, on the occasion of the 50th anniversary of this high school.

Jean-Jacques Lévy is member of the Inria-Alumni's executive committee (4 meetings in 2018) and organised the session about the Transparency of Algorithms (November 12).

Jean-Jacques Lévy was invited by the French Academy of Sciences to participate to the 2018 Hangzhou International Human Resources Exchanges and Cooperation Conference (Hangzhou, November 9-12).

Yann Régis-Gianas co-organised the "Journée Francilienne de Programmation", a programming contest between undergraduate students of three universities of Paris (UPD, UPMC, UPS).

### *8.3.1. Education*

Yann Régis-Gianas is the project leader of the "Learn-OCaml" project whose purpose is to support teaching the OCaml programming language worldwide.

### *8.3.2. Internal action*

- Science outreach towards services (DPEI, STIP...)

  Jean-Jacques Lévy talked about "L'informatique en 4 temps" at the Alumni-UniThé seminar at Inria Bordeaux (October 10).

# 9. Bibliography

## Major publications by the team in recent years

[1] R. M. AMADIO, Y. REGIS-GIANAS.*Certifying and reasoning about cost annotations of functional programs*, in "Higher-Order and Symbolic Computation", January 2013, https://hal.inria.fr/inria-00629473

[2] Z. ARIOLA, H. HERBELIN, A. SABRY.*A Type-Theoretic Foundation of Delimited Continuations*, in "Higher Order and Symbolic Computation", 2007, http://dx.doi.org/10.1007/s10990-007-9006-0

[3] D. BAELDE, A. DOUMANE, A. SAURIN.*Infinitary proof theory : the multiplicative additive case* , in "Proceedings of CSL 2016", September 2016, https://hal.archives-ouvertes.fr/hal-01339037

[4] C. CHENAVIER.*The lattice of reduction operators: applications to noncommutative Gröbner bases and homological algebra*, Université paris Diderot, December 2016, https://tel.archives-ouvertes.fr/tel-01415910

[5] P.-L. CURIEN.*Operads, clones, and distributive laws*, in "Operads and Universal Algebra : Proceedings of China-France Summer Conference", Tianjin, China, L. G. CHENGMING BAI, J.-L. LODAY (editors), Nankai Series in Pure, Applied Mathematics and Theoretical Physics, Vol. 9, World Scientific, July 2010, p. 25-50, https://hal.archives-ouvertes.fr/hal-00697065

[6] P.-L. CURIEN, R. GARNER, M. HOFMANN.*Revisiting the categorical interpretation of dependent type theory*, in "Theoretical computer Science", 2014, vol. 546, p. 99-119, http://dx.doi.org/10.1007/s10990-007-9006-0

[7] P.-L. CURIEN, H. HERBELIN.*The duality of computation*, in "Proceedings of the Fifth ACM SIGPLAN International Conference on Functional Programming (ICFP '00)", Montreal, Canada, SIGPLAN Notices 35(9), ACM, September 18-21 2000, p. 233–243 [*DOI :* 10.1145/351240.351262], http://hal.archives-ouvertes.fr/inria-00156377/en/

[8] P.-L. CURIEN, H. HERBELIN.*Abstract machines for dialogue games*, in "Interactive models of computation and program behavior", Panoramas et Synthèses, Société Mathématique de France, 2009, p. 231-275, https://hal.archives-ouvertes.fr/hal-00155295

[9] P. DEHORNOY, Y. GUIRAUD.*Quadratic normalization in monoids*, in "Internat. J. Algebra Comput.", 2016, vol. 26, n$^{o}$ 5, p. 935–972, https://doi.org/10.1142/S0218196716500399

[10] S. GAUSSENT, Y. GUIRAUD, P. MALBOS.*Coherent presentations of Artin monoids*, in "Compositio Mathematica", 2015, vol. 151, n$^{o}$ 5, p. 957-998 [*DOI :* 10.1112/S0010437X14007842], https://hal.archives-ouvertes.fr/hal-00682233

[11] G. GILBERT, J. COCKX, M. SOZEAU, N. TABAREAU.*Definitional Proof-Irrelevance without K*, in "46th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL 2019", Lisbon, Portugal, POPL, January 2019, https://hal.inria.fr/hal-01859964

[12] T. GIRKA, D. MENTRÉ, Y. REGIS-GIANAS.*Oracle-based Dierential Operational Semantics (long version)*, Université Paris Diderot / Sorbonne Paris Cité, October 2016, https://hal.inria.fr/hal-01419860

[13] Y. GUIRAUD.*Rewriting methods in homotopical and higher-dimensional algebra*, Univ. Paris 7, 2019, Habilitation à diriger des recherches

[14] Y. GUIRAUD, P. MALBOS.*Higher-dimensional normalisation strategies for acyclicity*, in "Advances in Mathematics", 2012, vol. 231, n$^{o}$ 3-4, p. 2294-2351 [*DOI :* 10.1016/J.AIM.2012.05.010], https://hal.archives-ouvertes.fr/hal-00531242

[15] Y. GUIRAUD, P. MALBOS, S. MIMRAM.*A Homotopical Completion Procedure with Applications to Coherence of Monoids*, in "RTA - 24th International Conference on Rewriting Techniques and Applications - 2013", Eindhoven, Netherlands, F. VAN RAAMSDONK (editor), Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, June 2013, vol. 21, p. 223-238 [*DOI :* 10.4230/LIPICS.RTA.2013.223], https://hal.inria.fr/hal-00818253

[16] H. HERBELIN.*On the Degeneracy of Sigma-Types in Presence of Computational Classical Logic*, in "Proceedings of TLCA 2005", P. URZYCZYN (editor), Lecture Notes in Computer Science, Springer, 2005, vol. 3461, p. 209–220

[17] H. HERBELIN.*An intuitionistic logic that proves Markov's principle*, in "Logic In Computer Science", Edinburgh, Royaume-Uni, IEEE Computer Society, 2010, http://hal.inria.fr/inria-00481815/en/

[18] H. HERBELIN.*A Constructive Proof of Dependent Choice, Compatible with Classical Logic*, in "LICS 2012 - 27th Annual ACM/IEEE Symposium on Logic in Computer Science", Dubrovnik, Croatia, Proceedings of the 27th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2012, 25-28 June 2012, Dubrovnik, Croatia, IEEE Computer Society, June 2012, p. 365-374, https://hal.inria.fr/hal-00697240

[19] G. JABER, N. TABAREAU, M. SOZEAU.*Extending Type Theory with Forcing*, in "LICS 2012 : Logic In Computer Science", Dubrovnik, Croatia, June 2012, https://hal.archives-ouvertes.fr/hal-00685150

[20] P. LETOUZEY.*Hofstadter's problem for curious readers*, Université Paris Diderot ; Inria Paris-Rocquencourt, September 2015, 29, https://hal.inria.fr/hal-01195587

[21] G. MUNCH-MACCAGNONI.*Focalisation and Classical Realisability*, in "Computer Science Logic '09", E. GRÄDEL, R. KAHLE (editors), Lecture Notes in Computer Science, Springer-Verlag, 2009, vol. 5771, p. 409–423

[22] T. U. F. PROGRAM.*Homotopy type theory—univalent foundations of mathematics*, The Univalent Foundations Program, Princeton, NJ; Institute for Advanced Study (IAS), Princeton, NJ, 2013, xiv+589, http://homotopytypetheory.org/book

[23] Y. REGIS-GIANAS, F. POTTIER.*A Hoare Logic for Call-by-Value Functional Programs*, in "Proceedings of the Ninth International Conference on Mathematics of Program Construction (MPC'08)", Lecture Notes in Computer Science, Springer, July 2008, vol. 5133, p. 305–335, http://gallium.inria.fr/~fpottier/publis/regis-gianas-pottier-hoarefp.ps.gz

[24] A. SAURIN.*Separation with Streams in the $\Lambda\mu$-calculus*, in "Symposium on Logic in Computer Science (LICS 2005)", Chicago, IL, USA, Proceedings, IEEE Computer Society, 26-29 June 2005, p. 356-365

[25] B. ZILIANI, M. SOZEAU.*A comprehensible guide to a new unifier for CIC including universe polymorphism and overloading*, in "Journal of Functional Programming", 2017, vol. 27 [*DOI :* 10.1017/S0956796817000028], https://hal.inria.fr/hal-01671925

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[26] G. CLARET.*Program in Coq*, Université Paris Diderot - Paris 7, September 2018, https://hal.inria.fr/tel-01890983

[27] T. GIRKA.*Differential program semantics*, Université Paris Diderot, July 2018, https://hal.inria.fr/tel-01890508

### Articles in International Peer-Reviewed Journal

[28] M. ANEL, G. BIEDERMANN, E. FINSTER, A. JOYAL.*Goodwillie's calculus of functors and higher topos theory*, in "Journal of topology", December 2018, vol. 11, n$^{o}$ 4, p. 1100-1132 [*DOI :* 10.1112/TOPO.12082], https://hal.inria.fr/hal-01939906

[29] F. FAGES, T. MARTINEZ, D. A. ROSENBLUETH, S. SOLIMAN.*Influence Networks compared with Reaction Networks: Semantics, Expressivity and Attractors*, in "IEEE/ACM Transactions on Computational Biology and Bioinformatics", 2018, vol. PP, n$^{o}$ 99, p. 1-14 [*DOI :* 10.1109/TCBB.2018.2805686], https://hal.inria.fr/hal-01510216

[30] G. GILBERT, J. COCKX, M. SOZEAU, N. TABAREAU.*Definitional Proof-Irrelevance without K*, in "Proceedings of the ACM on Programming Languages", January 2019, p. 1-28 [*DOI :* 10.1145/329031610.1145/3290316], https://hal.inria.fr/hal-01859964

[31] Y. GUIRAUD, E. HOFFBECK, P. MALBOS.*Convergent presentations and polygraphic resolutions of associative algebras*, in "Mathematische Zeitschrift", 2018, 68 pages, https://hal.archives-ouvertes.fr/hal-01006220

[32] Y. GUIRAUD, P. MALBOS.*Polygraphs of finite derivation type*, in "Mathematical Structures in Computer Science", 2018, vol. 28, n$^{o}$ 2, p. 155-201, https://arxiv.org/abs/1402.2587 [*DOI :* 10.1017/S0960129516000220], https://hal.archives-ouvertes.fr/hal-00932845

[33] Y. JIANG, J. LIU, G. DOWEK, K. JI.*Towards Combining Model Checking and Proof Checking*, in "The Computer Journal", 2019, https://hal.inria.fr/hal-01970274

[34] J.-O. KAISER, B. ZILIANI, R. KREBBERS, Y. RÉGIS-GIANAS, D. DREYER.*Mtac2: typed tactics for backward reasoning in Coq*, in "Proceedings of the ACM on Programming Languages", July 2018, vol. 2, n$^{o}$ ICFP, p. 1 - 31 [*DOI :* 10.1145/3236773], https://hal.inria.fr/hal-01890511

[35] L. PATEY, K. YOKOYAMA.*The proof-theoretic strength of Ramsey's theorem for pairs and two colors*, in "Advances in Mathematics", May 2018, vol. 330, p. 1034 - 1070 [*DOI :* 10.1016/J.AIM.2018.03.035], https://hal.archives-ouvertes.fr/hal-01888655

[36] N. TABAREAU, É. TANTER, M. SOZEAU.*Equivalences for Free: Univalent Parametricity for Effective Transport*, in "Proceedings of the ACM on Programming Languages", September 2018, p. 1-29 [*DOI :* 10.1145/3234615], https://hal.inria.fr/hal-01559073

### International Conferences with Proceedings

[37] A. ANAND, S. BOULIER, C. COHEN, M. SOZEAU, N. TABAREAU.*Towards Certified Meta-Programming with Typed Template-Coq*, in "ITP 2018 - 9th Conference on Interactive Theorem Proving", Oxford, United Kingdom, LNCS, Springer, July 2018, vol. 10895, p. 20-39 [*DOI :* 10.1007/978-3-319-94821-8_2], https://hal.archives-ouvertes.fr/hal-01809681

[38] T. LETAN, Y. RÉGIS-GIANAS, P. CHIFFLIER, G. HIET.*Modular Verification of Programs with Effects and Effect Handlers in Coq*, in "FM 2018 - 22nd International Symposium on Formal Methods", Oxford, United Kingdom, LNCS, Springer, July 2018, vol. 10951, p. 338-354 [*DOI :* 10.1007/978-3-319-95582-7_20], https://hal.inria.fr/hal-01799712

[39] É. MIQUEY, H. HERBELIN.*Realizability Interpretation and Normalization of Typed Call-by-Need λ-calculus With Control*, in "FOSSACS 18 - 21st International Conference on Foundations of Software Science and Computation Structures", Thessalonique, Greece, C. BAIER, U. D. LAGO (editors), LNCS, Springer, April 2018, vol. 10803, p. 276-292, https://arxiv.org/abs/1803.00914 [*DOI :* 10.1007/978-3-319-89366-2_15], https://hal.inria.fr/hal-01624839

[40] E. RIVAS.*Relating Idioms, Arrows and Monads from Monoidal Adjunctions*, in "Seventh Workshop on Mathematically Structured Functional Programming- EPTCS", Oxford, United Kingdom, July 2018, vol. 275, p. 18-33, https://hal.inria.fr/hal-01946996

[41] Y. RÉGIS-GIANAS, N. JEANNEROD, R. TREINEN.*Morbig: A Static Parser for POSIX Shell*, in "SLE 2018 - ACM SIGPLAN International Conference on Software Language Engineering", Boston, United States, November 2018 [*DOI :* 10.1145/3276604.3276615], https://hal.archives-ouvertes.fr/hal-01890044

[42] A. TIMANY, M. SOZEAU.*Cumulative Inductive Types in Coq*, in "FSCD 2018 - 3rd International Conference on Formal Structures for Computation and Deduction", Oxford, United Kingdom, July 2018 [*DOI :* 10.4230/LIPICS.FSCD.2018.29], https://hal.inria.fr/hal-01952037

[43] T. WINTERHALTER, M. SOZEAU, N. TABAREAU.*Eliminating Reflection from Type Theory: To the Legacy of Martin Hofmann*, in "CPP 2019 - The 8th ACM SIGPLAN International Conference on Certified Programs and Proofs", Lisbonne, Portugal, ACM, January 2019, p. 91-103 [*DOI :* 10.1145/3293880.3294095], https://hal.archives-ouvertes.fr/hal-01849166

### National Conferences with Proceeding

[44] P. LAFORGUE, Y. RÉGIS-GIANAS.*OCaml étendu avec du filtrage par comotifs*, in "JFLA 2018 - Journées Francophones des Langages Applicatifs", Banyuls sur mer, France, January 2018, https://hal.inria.fr/hal-01897456

### Conferences without Proceedings

[45] A. ANAND, S. BOULIER, N. TABAREAU, M. SOZEAU.*Typed Template Coq – Certified Meta-Programming in Coq*, in "CoqPL 2018 - The Fourth International Workshop on Coq for Programming Languages", Los Angeles, CA, United States, January 2018, p. 1-2, https://hal.inria.fr/hal-01671948

[46] R. NOLLET, A. SAURIN, C. TASSON.*Local validity for circular proofs in linear logic with fixed points: extended version*, in "Computer Science Logic", Birmingham, United Kingdom, September 2018, https://hal.archives-ouvertes.fr/hal-01825477

[47] T. ZIMMERMANN.*Challenges in the collaborative development of a complex mathematical software and its ecosystem*, in "OpenSym 2018 - 14th International Symposium on Open Collaboration", Paris, France, August 2018, vol. 2018 [*DOI :* 10.1145/3233391.3233966], https://hal.inria.fr/hal-01951322

### Scientific Books (or Scientific Book chapters)

[48] A. BOUAJJANI, C. ENEA, K. JI, S. QADEER.*On the Completeness of Verifying Message Passing Programs Under Bounded Asynchrony*, in "International Conference on Computer Aided Verification, CAV 2018: Computer Aided Verification", Springer International Publishing, July 2018, p. 372-391, https://hal.archives-ouvertes.fr/hal-01947855

### Other Publications

[49] R. CHEN, C. COHEN, J.-J. LEVY, S. MERZ, L. THERY.*Formal Proofs of Tarjan's Algorithm in Why3, Coq, and Isabelle*, October 2018, https://arxiv.org/abs/1810.11979 - working paper or preprint, https://hal.inria.fr/hal-01906155

[50] C. CHENAVIER.*A Lattice Formulation of the F 4 Completion Procedure*, January 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01489200

[51] C. CHENAVIER.*Syzygies among reduction operators*, April 2018, https://arxiv.org/abs/1708.08709 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01578555

[52] T. COQ DEVELOPMENT TEAM.*The Coq Proof Assistant, version 8.8.0*, April 2018, Software [*DOI :* 10.5281/ZENODO.1219885], https://hal.inria.fr/hal-01954564

[53] P.-L. CURIEN, J. OBRADOVIC.*Categorified cyclic operads*, January 2018, https://arxiv.org/abs/1706.06788 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01679682

[54] T. GIRKA.*correlating_program*, July 2018, https://archive.softwareheritage.org/swh:1:rev:c8fca417ee9eefe25683042192da67470 Software, https://hal.inria.fr/hal-01831364

[55] T. GIRKA, Y. RÉGIS-GIANAS.*Correlating Oracles*, July 2018, https://archive.softwareheritage.org/swh:1:rev:cccf789c12617208f Software, https://hal.inria.fr/hal-01831369

[56] H. HERBELIN, É. MIQUEY.*Continuation-and-environment-passing style translations: a focus on call-by-need*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01972846

[57] C. HO THANH.*The equivalence between many-to-one polygraphs and opetopic sets*, July 2018, https://arxiv.org/abs/1806.08645 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01946918

[58] C. MANGIN, M. SOZEAU.*Equations reloaded*, July 2018, working paper or preprint, https://hal.inria.fr/hal-01671777

[59] Y. REGIS-GIANAS, N. JEANNEROD, R. TREINEN.*Morbig*, October 2018, https://archive.softwareheritage.org/swh:1:dir:eb7770eb Software, https://hal.inria.fr/hal-01897572

[60] T. ZIMMERMANN, A. CASANUEVA ARTÍS.*Impact of switching bug trackers: a case study*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01951176

## References in notes

[61] T. ALTENKIRCH.*Extensional Equality in Intensional Type Theory*, in "14th Annual IEEE Symposium on Logic in Computer Science, Trento, Italy, July 2-5, 1999", IEEE Computer Society, 1999, p. 412–420, https://doi.org/10.1109/LICS.1999.782636

[62] T. ALTENKIRCH, P. CAPRIOTTI, N. KRAUS.*Extending Homotopy Type Theory with Strict Equality*, in "CSL", LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016, vol. 62, p. 21:1–21:17

[63] D. J. ANICK.*On the Homology of Associative Algebras*, in "Trans. Amer. Math. Soc.", 1986, vol. 296, n$^{o}$ 2, p. 641–659

[64] D. ARA, F. MÉTAYER.*The Brown-Golasiński Model Structure on strict $\infty$-groupoids revisited*, in "Homology, Homotopy and Applications", 2011, vol. 13, n$^{o}$ 1, p. 121–142

[65] J. BAEZ, A. CRANS.*Higher-dimensional algebra. VI. Lie 2-algebras*, in "Theory Appl. Categ.", 2004, vol. 12, p. 492–538

[66] H. P. BARENDREGT.*The Lambda Calculus: Its Syntax and Semantics*, North Holland, Amsterdam, 1984

[67] Y. BERTOT, P. CASTÉRAN.*Interactive Theorem Proving and Program Development Coq'Art: The Calculus of Inductive Constructions*, Springer, 2004

[68] G. BONFANTE, Y. GUIRAUD.*Polygraphic Programs and Polynomial-Time Functions*, in "Logical Methods in Computer Science", 2009, vol. 5, n$^{o}$ 2, p. 1–37

[69] S. BOULIER, P.-M. PÉDROT, N. TABAREAU.*The next 700 syntactical models of type theory*, in "Certified Programs and Proofs (CPP 2017)", Paris, France, January 2017, p. 182 - 194 [*DOI :* 10.1145/3018610.3018620], https://hal.inria.fr/hal-01445835

[70] B. BUCHBERGER.*Ein Algorithmus zum Auffinden der Basiselemente des Restklassenringes nach einem nulldimensionalen Polynomideal (An Algorithm for Finding the Basis Elements in the Residue Class Ring Modulo a Zero Dimensional Polynomial Ideal)*, Mathematical Institute, University of Innsbruck, Austria, 1965

[71] A. BURRONI.*Higher-dimensional word problems with applications to equational logic*, in "Theoretical Computer Science", jul 1993, vol. 115, n$^{o}$ 1, p. 43–62

[72] A. CHLIPALA.*Certified Programming with Dependent Types - A Pragmatic Introduction to the Coq Proof Assistant*, MIT Press, 2013, http://mitpress.mit.edu/books/certified-programming-dependent-types

[73] A. CHURCH.*A set of Postulates for the foundation of Logic*, in "Annals of Mathematics", 1932, vol. 2, p. 33, 346-366

[74] J. COCKX.*Dependent Pattern Matching and Proof-Relevant Unification*, Katholieke Universiteit Leuven, Belgium, 2017, https://lirias.kuleuven.be/handle/123456789/583556

[75] T. COQUAND.*Une théorie des Constructions*, University Paris 7, January 1985

[76] T. COQUAND, G. HUET.*Constructions : A Higher Order Proof System for Mechanizing Mathematics*, in "EUROCAL'85", Linz, Lecture Notes in Computer Science, Springer Verlag, 1985, vol. 203

[77] T. COQUAND, C. PAULIN-MOHRING.*Inductively defined types*, in "Proceedings of Colog'88", P. MARTIN-LÖF, G. MINTS (editors), Lecture Notes in Computer Science, Springer Verlag, 1990, vol. 417

[78] H. B. CURRY, R. FEYS, W. CRAIG.*Combinatory Logic*, North-Holland, 1958, vol. 1, §9E

[79] P. DEHORNOY, Y. LAFONT.*Homology of Gaussian groups*, in "Ann. Inst. Fourier (Grenoble)", 2003, vol. 53, n$^{o}$ 2, p. 489–540, http://aif.cedram.org/item?id=AIF_2003__53_2_489_0

[80] P. DELIGNE.*Action du groupe des tresses sur une catégorie*, in "Invent. Math.", 1997, vol. 128, n$^{o}$ 1, p. 159–175

[81] M. FELLEISEN, D. P. FRIEDMAN, E. KOHLBECKER, B. F. DUBA.*Reasoning with continuations*, in "First Symposium on Logic and Computer Science", 1986, p. 131-141

[82] A. FILINSKI.*Representing Monads*, in "Conf. Record 21st ACM SIGPLAN-SIGACT Symp. on Principles of Programming Languages, POPL'94", Portland, OR, USA, ACM Press, 17-21 Jan 1994, p. 446-457

[83] G. GENTZEN.*Untersuchungen über das logische Schließen*, in "Mathematische Zeitschrift", 1935, vol. 39, p. 176–210,405–431

[84] J.-Y. GIRARD.*Une extension de l'interpretation de Gödel à l'analyse, et son application à l'élimination des coupures dans l'analyse et la théorie des types*, in "Second Scandinavian Logic Symposium", J. FENSTAD (editor), Studies in Logic and the Foundations of Mathematics, North Holland, 1971, n$^{o}$ 63, p. 63-92

[85] T. G. GRIFFIN.*The Formulae-as-Types Notion of Control*, in "Conf. Record 17th Annual ACM Symp. on Principles of Programming Languages, POPL '90", San Francisco, CA, USA, 17-19 Jan 1990, ACM Press, 1990, p. 47–57

[86] Y. GUIRAUD.*Présentations d'opérades et systèmes de réécriture*, Univ. Montpellier 2, 2004

[87] Y. GUIRAUD.*Termination Orders for 3-Dimensional Rewriting*, in "Journal of Pure and Applied Algebra", 2006, vol. 207, n$^{o}$ 2, p. 341–371

[88] Y. GUIRAUD.*The Three Dimensions of Proofs*, in "Annals of Pure and Applied Logic", 2006, vol. 141, n$^{o}$ 1–2, p. 266–295

[89] Y. GUIRAUD.*Two Polygraphic Presentations of Petri Nets*, in "Theoretical Computer Science", 2006, vol. 360, n$^{o}$ 1–3, p. 124–146

[90] Y. GUIRAUD, E. HOFFBECK, P. MALBOS.*Confluence of linear rewriting and homology of algebras*, in "3rd International Workshop on Confluence", Vienna, Austria, July 2014, https://hal.archives-ouvertes.fr/hal-01105087

[91] Y. GUIRAUD, P. MALBOS.*Higher-dimensional categories with finite derivation type*, in "Theory Appl. Categ.", 2009, vol. 22, n$^{o}$ 18, p. 420-478

[92] Y. GUIRAUD, P. MALBOS.*Identities among relations for higher-dimensional rewriting systems*, in "Séminaires et Congrès, Société Mathématique de France", 2011, vol. 26, p. 145-161

[93] Y. GUIRAUD, P. MALBOS.*Coherence in monoidal track categories*, in "Math. Structures Comput. Sci.", 2012, vol. 22, n$^{o}$ 6, p. 931–969

[94] M. HOFMANN, T. STREICHER.*The groupoid interpretation of type theory*, in "Twenty-five years of constructive type theory (Venice, 1995)", Oxford Logic Guides, Oxford Univ. Press, New York, 1998, vol. 36, p. 83–111

[95] W. A. HOWARD.*The formulae-as-types notion of constructions*, in "to H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism", Academic Press, 1980, Unpublished manuscript of 1969

[96] J. KOCK, A. JOYAL, M. BATANIN, J.-F. MASCARI.*Polynomial functors and opetopes*, in "Advances in Mathematics", 2010, vol. 224, n° 6, p. 2690 - 2737 [*DOI :* 10.1016/J.AIM.2010.02.012], http://www.sciencedirect.com/science/article/pii/S0001870810000769

[97] J.-L. KRIVINE.*A call-by-name lambda-calculus machine*, in "Higher Order and Symbolic Computation", 2005

[98] J.-L. KRIVINE.*Un interpréteur du lambda-calcul*, 1986, Unpublished

[99] Y. LAFONT.*Towards an Algebraic Theory of Boolean Circuits*, in "Journal of Pure and Applied Algebra", 2003, vol. 184, p. 257-310

[100] Y. LAFONT, F. MÉTAYER, K. WORYTKIEWICZ.*A Folk Model Structure on Omega-Cat*, in "Advances in Mathematics", 2010, vol. 224, n° 3, p. 1183–1231

[101] P. LANDIN.*The mechanical evaluation of expressions*, in "The Computer Journal", January 1964, vol. 6, n° 4, p. 308–320

[102] P. LANDIN.*A generalisation of jumps and labels*, UNIVAC Systems Programming Research, August 1965, n° ECS-LFCS-88-66, Reprinted in Higher Order and Symbolic Computation, 11(2), 1998

[103] J. LURIE.*Higher topos theory*, Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, 2009, vol. 170, xviii+925

[104] P. MALBOS.*Critères de finitude homologique pour la non convergence des systèmes de réécriture de termes*, Univ. Montpellier 2, 2004

[105] P. MARTIN-LÖF.*A theory of types*, University of Stockholm, 1971, n° 71-3

[106] M. PARIGOT.*Free Deduction: An Analysis of "Computations" in Classical Logic*, in "Logic Programming, Second Russian Conference on Logic Programming", St. Petersburg, Russia, A. VORONKOV (editor), Lecture Notes in Computer Science, Springer, September 11-16 1991, vol. 592, p. 361-380, http://www.informatik.uni-trier.de/~ley/pers/hd/p/Parigot:Michel.html

[107] S. B. PRIDDY.*Koszul resolutions*, in "Trans. Amer. Math. Soc.", 1970, vol. 152, p. 39–60

[108] J. C. REYNOLDS.*Definitional interpreters for higher-order programming languages*, in "ACM '72: Proceedings of the ACM annual conference", New York, NY, USA, ACM Press, 1972, p. 717–740

[109] J. C. REYNOLDS.*Towards a theory of type structure*, in "Symposium on Programming", B. ROBINET (editor), Lecture Notes in Computer Science, Springer, 1974, vol. 19, p. 408-423

[110] C. SQUIER, F. OTTO, Y. KOBAYASHI.*A finiteness condition for rewriting systems*, in "Theoret. Comput. Sci.", 1994, vol. 131, n$^{o}$ 2, p. 271–294

[111] C. C. SQUIER.*Word problems and a homological finiteness condition for monoids*, in "J. Pure Appl. Algebra", 1987, vol. 49, n$^{o}$ 1-2, p. 201–217

[112] R. STREET.*Limits Indexed by Category-Valued 2-Functors*, in "Journal of Pure and Applied Algebra", 1976, vol. 8, p. 149–181

[113] T. C. D. TEAM.*The Coq Proof Assistant, version 8.7.1*, December 2017, https://doi.org/10.5281/zenodo.1133970

[114] N. DE BRUIJN.*AUTOMATH, a language for mathematics*, Technological University Eindhoven, November 1968, n$^{o}$ 66-WSK-05

# Project-Team POLSYS

## Polynomial Systems

IN COLLABORATION WITH: Laboratoire d'informatique de Paris 6 (LIP6)

IN PARTNERSHIP WITH:
**CNRS**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Algorithmics, Computer Algebra and Cryptology**

# Table of contents

# Project-Team POLSYS

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

**Keywords:**

**Computer Science and Digital Science:**

A2.4. - Formal method for verification, reliability, certification
A4.3. - Cryptography
A4.3.1. - Public key cryptography
A4.3.4. - Quantum Cryptography
A5.10.1. - Design
A6.1. - Methods in mathematical modeling
A6.2.3. - Probabilistic methods
A6.2.6. - Optimization
A6.2.7. - High performance computing
A6.4.3. - Observability and Controlability
A8.1. - Discrete mathematics, combinatorics
A8.2. - Optimization
A8.3. - Geometry, Topology
A8.4. - Computer Algebra

**Other Research Topics and Application Domains:**

B5. - Industry of the future
B5.2. - Design and manufacturing
B5.2.3. - Aviation
B5.2.4. - Aerospace
B6. - IT and telecom
B6.3. - Network functions
B6.5. - Information systems
B9.5.1. - Computer science
B9.5.2. - Mathematics
B9.10. - Privacy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Jean-Charles Faugère [Team leader, Inria, Senior Researcher, HDR]
Elias Tsigaridas [Inria, Researcher]
Dongming Wang [CNRS, Senior Researcher, on leave at Beihang University, HDR]

**Faculty Members**

Jérémy Berthomieu [Sorbonne Université, Associate Professor]
Daniel Lazard [Sorbonne Université, Emeritus Professor, HDR]
Ludovic Perret [Sorbonne Université, Associate Professor, HDR]
Guénaël Renault [Sorbonne Université, Associate Professor, on leave at ANSSI, HDR]
Mohab Safey El Din [Sorbonne Université, Professor, HDR]

**External Collaborators**

Emmanuel Prouff [ANSSI, Associate Member, HDR]
Victor Magron [CNRS, Researcher, until Aug. 2018]

**PhD Students**

Matías Bender [Inria, until Jun. 2019]
Olive Chakraborty [Sorbonne Université]
Nagardjun Chinthamani Dwarakanath [Sorbonne Université]
Solane El Hirch [Sorbonne Université]
Phuoc Le [Sorbonne Université]
Jocelyn Ryckeghem [Sorbonne Université]
Thi Xuan Vu [Sorbonne Université]

**Post-Doctoral Fellows**

Rachel Player [Sorbonne Université, Post-Doctoral fellow, until Dec. 2018]
Kaie Kubjas [Sorbonne Université, Post-Doctoral fellow]
Amine Mrabet [Sorbonne Université, ATER, until Aug. 2018]

**Administrative Assistants**

Christelle Guiziou [Inria, from Nov. 2018]
Irphane Khan [Sorbonne Université, Assistant]

# 2. Overall Objectives

## 2.1. Overall Objectives

The main focus of the POLSYS project is to solve systems of polynomial equations.

Our main objectives are:

- **Fundamental Algorithms and Structured Systems.** The objective is to propose fast exponential exact algorithms for solving polynomial equations and to identify large classes of structured polynomial systems which can be solved in polynomial time.
- **Solving Systems over the Reals and Applications.** For positive dimensional systems basic questions over the reals may be very difficult (for instance testing the existence of solutions) but also very useful in applications (e.g. global optimization problems). We plan to propose efficient algorithms and implementations to address the most important issues: computing sample points in the real solution sets, decide if two such sample points can be path-connected and, as a long term objective, perform quantifier elimination over the reals (computing a quantifier-free formula which is equivalent to a given quantified boolean formula of polynomial equations/inequalities).
- **Dedicated Algebraic Computation and Linear Algebra.** While linear algebra is a key step in the computation of Gröbner bases, the matrices generated by the algorithms $F_4/F_5$ have specific structures (quasi block triangular). The objective is to develop a dedicated efficient multi-core linear algebra package as the basis of a future open source library for computing Gröbner bases.
- **Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory.** We propose to develop a systematic use of *structured systems* in Algebraic Cryptanalysis. We want to improve the efficiency and to predict the theoretical complexity of such attacks. We plan to demonstrate the power of algebraic techniques in new areas of cryptography such as Algebraic Number Theory (typically, in curve based cryptography).

# 3. Research Program

## 3.1. Introduction

Polynomial system solving is a fundamental problem in Computer Algebra with many applications in cryptography, robotics, biology, error correcting codes, signal theory, ... Among all available methods for solving polynomial systems, computation of Gröbner bases remains one of the most powerful and versatile method since it can be applied in the continuous case (rational coefficients) as well as in the discrete case (finite fields). Gröbner bases are also building blocks for higher level algorithms that compute real sample points in the solution set of polynomial systems, decide connectivity queries and quantifier elimination over the reals. The major challenge facing the designer or the user of such algorithms is the intrinsic exponential behaviour of the complexity for computing Gröbner bases. The current proposal is an attempt to tackle these issues in a number of different ways: improve the efficiency of the fundamental algorithms (even when the complexity is exponential), develop high performance implementation exploiting parallel computers, and investigate new classes of structured algebraic problems where the complexity drops to polynomial time.

## 3.2. Fundamental Algorithms and Structured Systems

**Participants:** Jérémy Berthomieu, Jean-Charles Faugère, Guénaël Renault, Mohab Safey El Din, Elias Tsigaridas, Dongming Wang, Matías Bender, Thi Xuan Vu.

Efficient algorithms $F_4/F_5$[0] for computing the Gröbner basis of a polynomial system rely heavily on a connection with linear algebra. Indeed, these algorithms reduce the Gröbner basis computation to a sequence of Gaussian eliminations on several submatrices of the so-called Macaulay matrix in some degree. Thus, we expect to improve the existing algorithms by
*(i)* developing dedicated linear algebra routines performing the Gaussian elimination steps: this is precisely the objective 2 described below;
*(ii)* generating smaller or simpler matrices to which we will apply Gaussian elimination.
We describe here our goals for the latter problem. First, we focus on algorithms for computing a Gröbner basis of *general polynomial systems*. Next, we present our goals on the development of dedicated algorithms for computing Gröbner bases of *structured polynomial systems* which arise in various applications.

**Algorithms for general systems.** Several degrees of freedom are available to the designer of a Gröbner basis algorithm to generate the matrices occurring during the computation. For instance, it would be desirable to obtain matrices which would be almost triangular or very sparse. Such a goal can be achieved by considering various interpretations of the $F_5$ algorithm with respect to different monomial orderings. To address this problem, the tight complexity results obtained for $F_5$ will be used to help in the design of such a general algorithm. To illustrate this point, consider the important problem of solving boolean polynomial systems; it might be interesting to preserve the sparsity of the original equations and, at the same time, using the fact that overdetermined systems are much easier to solve.

**Algorithms dedicated to *structured* polynomial systems.** A complementary approach is to exploit the structure of the input polynomials to design specific algorithms. Very often, problems coming from applications are not random but are highly structured. The specific nature of these systems may vary a lot: some polynomial systems can be sparse (when the number of terms in each equation is low), overdetermined (the number of the equations is larger than the number of variables), invariants by the action of some finite groups, multi-linear (each equation is linear w.r.t. to one block of variables) or more generally multihomogeneous. In each case, the ultimate goal is to identify large classes of problems whose theoretical/practical complexity drops and to propose in each case dedicated algorithms.

[0] J.-C. Faugère. *A new efficient algorithm for computing Gröbner bases without reduction to zero (F5).* In Proceedings of ISSAC '02, pages 75-83, New York, NY, USA, 2002. ACM.

## 3.3. Solving Systems over the Reals and Applications.

**Participants:** Mohab Safey El Din, Elias Tsigaridas, Daniel Lazard, Ivan Bannwarth, Thi Xuan Vu.

We shall develop algorithms for solving polynomial systems over complex/real numbers. Again, the goal is to extend significantly the range of reachable applications using algebraic techniques based on Gröbner bases and dedicated linear algebra routines. Targeted application domains are global optimization problems, stability of dynamical systems (e.g. arising in biology or in control theory) and theorem proving in computational geometry.

The following functionalities shall be requested by the end-users:
*(i)* deciding the emptiness of the real solution set of systems of polynomial equations and inequalities,
*(ii)* quantifier elimination over the reals or complex numbers,
*(iii)* answering connectivity queries for such real solution sets.
We will focus on these functionalities.

We will develop algorithms based on the so-called critical point method to tackle systems of equations and inequalities (problem *(i)*) . These techniques are based on solving 0-dimensional polynomial systems encoding "critical points" which are defined by the vanishing of minors of Jacobian matrices (with polynomial entries). Since these systems are highly structured, the expected results of Objective 1 and 2 may allow us to obtain dramatic improvements in the computation of Gröbner bases of such polynomial systems. This will be the foundation of practically fast implementations (based on singly exponential algorithms) outperforming the current ones based on the historical Cylindrical Algebraic Decomposition (CAD) algorithm (whose complexity is doubly exponential in the number of variables). We will also develop algorithms and implementations that allow us to analyze, at least locally, the topology of solution sets in some specific situations. A long-term goal is obviously to obtain an analysis of the global topology.

## 3.4. Low level implementation and Dedicated Algebraic Computation and Linear Algebra.

**Participants:** Jean-Charles Faugère, Elias Tsigaridas, Olive Chakraborty, Jocelyn Ryckeghem.

Here, the primary objective is to focus on *dedicated* algorithms and software for the linear algebra steps in Gröbner bases computations and for problems arising in Number Theory. As explained above, linear algebra is a key step in the process of computing efficiently Gröbner bases. It is then natural to develop specific linear algebra algorithms and implementations to further strengthen the existing software. Conversely, Gröbner bases computation is often a key ingredient in higher level algorithms from Algebraic Number Theory. In these cases, the algebraic problems are very particular and specific. Hence dedicated Gröbner bases algorithms and implementations would provide a better efficiency.

**Dedicated linear algebra tools.** The FGB library is an efficient one for Gröbner bases computations which can be used, for instance, via MAPLE. However, the library is sequential. A goal of the project is to extend its efficiency to new trend parallel architectures such as clusters of multi-processor systems in order to tackle a broader class of problems for several applications. Consequently, our first aim is to provide a durable, long term software solution, which will be the successor of the existing FGB library. To achieve this goal, we will first develop a high performance linear algebra package (under the LGPL license). This could be organized in the form of a collaborative project between the members of the team. The objective is not to develop a general library similar to the LINBOX[0] project but to propose a dedicated linear algebra package taking into account the specific properties of the matrices generated by the Gröbner bases algorithms. Indeed these matrices are sparse (the actual sparsity depends strongly on the application), almost block triangular and not necessarily of full rank. Moreover, most of the pivots are known at the beginning of the computation. In practice, such matrices are huge (more than $10^6$ columns) but taking into account their shape may allow us to speed up the computations by one or several orders of magnitude. A variant of a Gaussian elimination algorithm together with a corresponding C implementation has been presented. The main peculiarity is the order in which the operations are performed. This will be the kernel of the new linear algebra library that will be developed.

[0] http://www.linalg.org/

Fast linear algebra packages would also benefit to the transformation of a Gröbner basis of a zero–dimensional ideal with respect to a given monomial ordering into a Gröbner basis with respect to another ordering. In the generic case at least, the change of ordering is equivalent to the computation of the minimal polynomial of a so-called multiplication matrix. By taking into account the sparsity of this matrix, the computation of the Gröbner basis can be done more efficiently using a variant of the Wiedemann algorithm. Hence, our goal is also to obtain a dedicated high performance library for transforming (i.e. change ordering) Gröbner bases.

**Dedicated algebraic tools for Algebraic Number Theory.** Recent results in Algebraic Number Theory tend to show that the computation of Gröbner basis is a key step toward the resolution of difficult problems in this domain [0]. Using existing resolution methods is simply not enough to solve relevant problems. The main algorithmic bottleneck to overcome is to adapt the Gröbner basis computation step to the specific problems. Typically, problems coming from Algebraic Number Theory usually have a lot of symmetries or the input systems are very structured. This is the case, in particular, for problems coming from the algorithmic theory of Abelian varieties over finite fields [0] where the objects are represented by polynomial system and are endowed with intrinsic group actions. The main goal here is to provide dedicated algebraic resolution algorithms and implementations for solving such problems. We do not restrict our focus on problems in positive characteristic. For instance, tower of algebraic fields can be viewed as triangular sets; more generally, related problems (e.g. effective Galois theory) which can be represented by polynomial systems will receive our attention. This is motivated by the fact that, for example, computing small integer solutions of Diophantine polynomial systems in connection with Coppersmith's method would also gain in efficiency by using a dedicated Gröbner bases computations step.

## 3.5. Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory.

**Participants:** Jérémy Berthomieu, Jean-Charles Faugère, Ludovic Perret, Guénaël Renault, Olive Chakraborty, Nagardjun Chinthamani, Solane El Hirch, Jocelyn Ryckeghem.

Here, we focus on solving polynomial systems over finite fields (i.e. the discrete case) and the corresponding applications (Cryptology, Error Correcting Codes, ...). Obviously this objective can be seen as an application of the results of the two previous objectives. However, we would like to emphasize that it is also the source of new theoretical problems and practical challenges. We propose to develop a systematic use of *structured systems* in *algebraic cryptanalysis*.
*(i)* So far, breaking a cryptosystem using algebraic techniques could be summarized as modeling the problem by algebraic equations and then computing a, usually, time consuming Gröbner basis. A new trend in this field is to require a theoretical complexity analysis. This is needed to explain the behavior of the attack but also to help the designers of new cryptosystems to propose actual secure parameters.
*(ii)* To assess the security of several cryptosystems in symmetric cryptography (block ciphers, hash functions, ...), a major difficulty is the size of the systems involved for this type of attack. More specifically, the bottleneck is the size of the linear algebra problems generated during a Gröbner basis computation.

We propose to develop a systematic use of *structured systems* in *algebraic cryptanalysis*.

[0] P. Gaudry, *Index calculus for abelian varieties of small dimension and the elliptic curve discrete logarithm problem*, Journal of Symbolic Computation 44,12 (2009) pp. 1690-1702
[0] e.g. point counting, discrete logarithm, isogeny.

The first objective is to build on the recent breakthrough in attacking McEliece's cryptosystem: it is the first structural weakness observed on one of the oldest public key cryptosystem. We plan to develop a well founded framework for assessing the security of public key cryptosystems based on coding theory from the algebraic cryptanalysis point of view. The answer to this issue is strongly related to the complexity of solving bihomogeneous systems (of bidegree $(1, d)$). We also plan to use the recently gained understanding on the complexity of structured systems in other areas of cryptography. For instance, the MinRank problem – which can be modeled as an overdetermined system of bilinear equations – is at the heart of the structural attack proposed by Kipnis and Shamir against HFE (one of the most well known multivariate public cryptosystem). The same family of structured systems arises in the algebraic cryptanalysis of the Discrete Logarithmic Problem (DLP) over curves (defined over some finite fields). More precisely, some bilinear systems appear in the polynomial modeling the points decomposition problem. Moreover, in this context, a natural group action can also be used during the resolution of the considered polynomial system.

Dedicated tools for linear algebra problems generated during the Gröbner basis computation will be used in algebraic cryptanalysis. The promise of considerable algebraic computing power beyond the capability of any standard computer algebra system will enable us to attack various cryptosystems or at least to propose accurate secure parameters for several important cryptosystems. Dedicated linear tools are thus needed to tackle these problems. From a theoretical perspective, we plan to further improve the theoretical complexity of the hybrid method and to investigate the problem of solving polynomial systems with noise, i.e. some equations of the system are incorrect. The hybrid method is a specific method for solving polynomial systems over finite fields. The idea is to mix exhaustive search and Gröbner basis computation to take advantage of the over-determinacy of the resulting systems.

Polynomial system with noise is currently emerging as a problem of major interest in cryptography. This problem is a key to further develop new applications of algebraic techniques; typically in side-channel and statistical attacks. We also emphasize that recently a connection has been established between several classical lattice problems (such as the Shortest Vector Problem), polynomial system solving and polynomial systems with noise. The main issue is that there is no sound algorithmic and theoretical framework for solving polynomial systems with noise. The development of such framework is a long-term objective.

# 4. Highlights of the Year

## 4.1. Highlights of the Year

Jean-Charles Faugère and Ludovic Perret received the Atos-Joseph Fourier 2018 prize [0] for their project on Quantum Safe Security.

# 5. New Software and Platforms

## 5.1. Epsilon

FUNCTIONAL DESCRIPTION: Epsilon is a library of functions implemented in Maple and Java for polynomial elimination and decomposition with (geometric) applications.

- Contact: Dongming Wang
- URL: http://wang.cc4cm.org/epsilon/index.html

## 5.2. FGb

KEYWORDS: Gröbner bases - Nonlinear system - Computer algebra

[0] https://atos.net/fr/2018/communiques-de-presse_2018_07_06/atos-et-genci-annoncent-les-laureats-du-prix-atos-joseph-fourier-2018

FUNCTIONAL DESCRIPTION: FGb is a powerful software for computing Gröbner bases. It includes the new generation of algorihms for computing Gröbner bases polynomial systems (mainly the F4, F5 and FGLM algorithms). It is implemented in C/C++ (approximately 250000 lines), standalone servers are available on demand. Since 2006, FGb is dynamically linked with Maple software (version 11 and higher) and is part of the official distribution of this software.

- Participant: Jean Charles Faugere
- Contact: Jean-Charles Faugère
- URL: http://www-polsys.lip6.fr/~jcf/FGb/index.html

## 5.3. FGb Light

FUNCTIONAL DESCRIPTION: Gröbner basis computation modulo p (p is a prime integer of 16 bits).

- Participant: Jean-Charles Faugère
- Contact: Jean-Charles Faugère
- URL: http://www-polsys.lip6.fr/~jcf/FGb/index.html

## 5.4. GBLA

FUNCTIONAL DESCRIPTION: GBLA is an open source C library for linear algebra specialized for eliminating matrices generated during Gröbner basis computations in algorithms like F4 or F5.

- Contact: Jean-Charles Faugère
- URL: http://www-polsys.lip6.fr/~jcf/GBLA/index.html

## 5.5. HFEBoost

FUNCTIONAL DESCRIPTION: Public-key cryptography system enabling an authentification of dematerialized data.

- Authors: Jean-Charles Faugère and Ludovic Perret
- Partner: UPMC
- Contact: Jean-Charles Faugère
- URL: http://www-polsys.lip6.fr/Links/hfeboost.html

## 5.6. RAGlib

*Real Algebraic Geometry library*
FUNCTIONAL DESCRIPTION: RAGLib is a powerful library, written in Maple, dedicated to solving over the reals polynomial systems. It is based on the FGb library for computing Grobner bases. It provides functionalities for deciding the emptiness and/or computing sample points to real solution sets of polynomial systems of equations and inequalities. This library provides implementations of the state-of-the-art algorithms with the currently best known asymptotic complexity for those problems.

- Contact: Mohab Safey El Din
- URL: http://www-polsys.lip6.fr/~safey/RAGLib/

## 5.7. RealCertify

KEYWORDS: Polynomial or analytical systems - Univariate polynomial - Real solving

FUNCTIONAL DESCRIPTION: The package RealCertify aims at providing a full suite of hybrid algorithms for computing certificates of non-negativity based on numerical software for solving linear matrix inequalities. The module univsos handles the univariate case and the module multivsos is designed for the multivariate case.

- Contact: Mohab Safey El Din
- URL: https://gricad-gitlab.univ-grenoble-alpes.fr/magronv/RealCertify

## 5.8. SLV

FUNCTIONAL DESCRIPTION: SLV is a software package in C that provides routines for isolating (and subsequently refine) the real roots of univariate polynomials with integer or rational coefficients based on subdivision algorithms and on the continued fraction expansion of real numbers. Special attention is given so that the package can handle polynomials that have degree several thousands and size of coefficients hundrends of Megabytes. Currently the code consists of approx. 5000 lines.

- Contact: Elias Tsigaridas
- URL: http://www-polsys.lip6.fr/~elias/soft

## 5.9. SPECTRA

*Semidefinite Programming solved Exactly with Computational Tools of Real Algebra*
KEYWORD: Linear Matrix Inequalities
FUNCTIONAL DESCRIPTION: SPECTRA is a Maple library devoted to solving exactly Semi-Definite Programs. It can handle rank constraints on the solution. It is based on the FGb library for computing Gröbner bases and provides either certified numerical approximations of the solutions or exact representations thereof.

- Contact: Mohab Safey El Din
- URL: http://homepages.laas.fr/henrion/software/spectra/

# 6. New Results

## 6.1. Fundamental algorithms and structured polynomial systems

### *6.1.1. Towards Mixed Gröbner Basis Algorithms: the Multihomogeneous and Sparse Case*

One of the biggest open problems in computational algebra is the design of efficient algorithms for Gröbner basis computations that take into account the sparsity of the input polynomials. We can perform such computations in the case of unmixed polynomial systems, that is systems with polynomials having the same support, using the approach of Faugère, Spaenlehauer, and Svartz [ISSAC'14]. In [15] we present two algorithms for sparse Gröbner bases computations for mixed systems. The first one computes with mixed sparse systems and exploits the supports of the polynomials. Under regularity assumptions, it performs no reductions to zero. For mixed, square, and 0-dimensional multihomogeneous polynomial systems, we present a dedicated, and potentially more efficient, algorithm that exploits different algebraic properties that performs no reduction to zero. We give an explicit bound for the maximal degree appearing in the computations.

### *6.1.2. Bilinear Systems with Two Supports: Koszul Resultant Matrices, Eigenvalues, and Eigenvectors*

A fundamental problem in computational algebraic geometry is the computation of the resultant. A central question is when and how to compute it as the determinant of a matrix whose elements are the coefficients of the input polynomials up-to sign. This problem is well understood for unmixed multihomogeneous systems, that is for systems consisting of multihomogeneous polynomials with the same support. However, little is known for mixed systems, that is for systems consisting of polynomials with different supports. In [14] we consider the computation of the multihomogeneous resultant of bilinear systems involving two different supports. We present a constructive approach that expresses the resultant as the exact determinant of a *Koszul resultant matrix*, that is a matrix constructed from maps in the Koszul complex. We exploit the resultant matrix to propose an algorithm to solve such systems. In the process we extend the classical eigenvalues and eigenvectors criterion to a more general setting. Our extension of the eigenvalues criterion applies to a general class of matrices, including the Sylvester-type and the Koszul-type ones.

### *6.1.3. A Polynomial-Division-Based Algorithm for Computing Linear Recurrence Relations*

Sparse polynomial interpolation, sparse linear system solving or modular rational reconstruction are fundamental problems in Computer Algebra. They come down to computing linear recurrence relations of a sequence with the Berlekamp–Massey algorithm. Likewise, sparse multivariate polynomial interpolation and multidimensional cyclic code decoding require guessing linear recurrence relations of a multivariate sequence.

Several algorithms solve this problem. The so-called Berlekamp–Massey–Sakata algorithm (1988) uses polynomial additions and shifts by a monomial. The SCALAR-FGLM algorithm (2015) relies on linear algebra operations on a multi-Hankel matrix, a multivariate generalization of a Hankel matrix. The Artinian Gorenstein border basis algorithm (2017) uses a Gram-Schmidt process.

In [16], we propose a new algorithm for computing the Gröbner basis of the ideal of relations of a sequence based solely on multivariate polynomial arithmetic. This algorithm allows us to both revisit the Berlekamp–Massey–Sakata algorithm through the use of polynomial divisions and to completely revise the SCALAR-FGLM algorithm without linear algebra operations.

A key observation in the design of this algorithm is to work on the mirror of the truncated generating series allowing us to use polynomial arithmetic modulo a monomial ideal. It appears to have some similarities with Padé approximants of this mirror polynomial.

Finally, we give a partial solution to the transformation of this algorithm into an adaptive one.

As an addition from the paper published at the ISSAC conferance, in [24], we give an adaptive variant of this algorithm taking into account the shape of the final Gröbner basis gradually as it is discovered. The main advantage of this algorithm is that its complexity in terms of operations and sequence queries only depends on the output Gröbner basis.

All these algorithms have been implemented in MAPLE and we report on our comparisons.

### *6.1.4. In-depth comparison of the Berlekamp–Massey–Sakata and the Scalar-FGLM algorithms: the adaptive variants*

The BERLEKAMP–MASSEY–SAKATA algorithm and the SCALAR-FGLM algorithm both compute the ideal of relations of a multidimensional linear recurrent sequence.

Whenever quering a single sequence element is prohibitive, the bottleneck of these algorithms becomes the computation of all the needed sequence terms. As such, having adaptive variants of these algorithms, reducing the number of sequence queries, becomes mandatory.

A native adaptive variant of the SCALAR-FGLM algorithm was presented by its authors, the so-called ADAPTIVE SCALAR-FGLM algorithm.

In [25], our first contribution is to make the BERLEKAMP–MASSEY–SAKATA algorithm more efficient by making it adaptive to avoid some useless relation testings. This variant allows us to divide by four in dimension 2 and by seven in dimension 3 the number of basic operations performed on some sequence family.

Then, we compare the two adaptive algorithms. We show that their behaviors differ in a way that it is not possible to tweak one of the algorithms in order to mimic exactly the behavior of the other. We detail precisely the differences and the similarities of both algorithms and conclude that in general the ADAPTIVE SCALAR-FGLM algorithm needs fewer queries and performs fewer basic operations than the ADAPTIVE BERLEKAMP–MASSEY–SAKATA algorithm.

We also show that these variants are always more efficient than the original algorithms.

### 6.1.5. Bit complexity for multi-homogeneous polynomial system solving Application to polynomial minimization

Multi-homogeneous polynomial systems arise in many applications. In [10] we provide bit complexity estimates for solving them which, up to a few extra other factors, are quadratic in the number of solutions and linear in the height of the input system under some genericity assumptions. The assumptions essentially imply that the Jacobian matrix of the system under study has maximal rank at the solution set and that this solution set if finite. The algorithm is probabilistic and a probability analysis is provided. Next, we apply these results to the problem of optimizing a linear map on the real trace of an algebraic set. Under some genericity assumptions, we provide bit complexity estimates for solving this polynomial minimization problem.

## 6.2. Solving Systems over the Reals and Applications

### 6.2.1. Univariate real root isolation in an extension field and applications

In [11] we present algorithmic, complexity and implementation results for the problem of isolating the real roots of a univariate polynomial in $B_\alpha \in L[y]$, where $L = \mathbb{Q}(\alpha)$ is a simple algebraic extension of the rational numbers. We revisit two approaches for the problem. In the first approach, using resultant computations, we perform a reduction to a polynomial with integer coefficients and we deduce a bound of $\widetilde{\mathcal{O}}_B(N^8)$ for isolating the real roots of $B_\alpha$, where $N$ is an upper bound on all the quantities (degree and bitsize) of the input polynomials. The bound becomes $\widetilde{\mathcal{O}}_B(N^7)$ if we use Pan's algorithm for isolating the real roots. In the second approach we isolate the real roots working directly on the polynomial of the input. We compute improved separation bounds for the roots and we prove that they are optimal, under mild assumptions. For isolating the real roots we consider a modified Sturm algorithm, and a modified version of `descartes`' algorithm. For the former we prove a Boolean complexity bound of $\widetilde{\mathcal{O}}_B(N^{12})$ and for the latter a bound of $\widetilde{\mathcal{O}}_B(N^5)$. We present aggregate separation bounds and complexity results for isolating the real roots of all polynomials $B_{\alpha_k}$, when $\alpha_k$ runs over all the real conjugates of $\alpha$. We show that we can isolate the real roots of all polynomials in $\widetilde{\mathcal{O}}_B(N^5)$. Finally, we implemented the algorithms in `C` as part of the core library of MATHEMATICA and we illustrate their efficiency over various data sets.

### 6.2.2. On the Maximal Number of Real Embeddings of Spatial Minimally Rigid Graphs

The number of embeddings of minimally rigid graphs in $\mathbb{R}^D$ is (by definition) finite, modulo rigid transformations, for every generic choice of edge lengths. Even though various approaches have been proposed to compute it, the gap between upper and lower bounds is still enormous. Specific values and its asymptotic behavior are major and fascinating open problems in rigidity theory. Our work in [13] considers the maximal number of real embeddings of minimally rigid graphs in $\mathbb{R}^3$. We modify a commonly used parametric semi-algebraic formulation that exploits the Cayley-Menger determinant to minimize the *a priori* number of complex embeddings, where the parameters correspond to edge lengths. To cope with the huge dimension of the parameter space and find specializations of the parameters that maximize the number of real embeddings, we introduce a method based on coupler curves that makes the sampling feasible for spatial minimally rigid graphs. Our methodology results in the first full classification of the number of real embeddings of graphs with 7 vertices in $\mathbb{R}^3$, which was the smallest open case. Building on this and certain 8-vertex graphs, we improve the previously known general lower bound on the maximum number of real embeddings in $\mathbb{R}^3$.

### *6.2.3. Lower bounds on the number of realizations of rigid graphs*

Computing the number of realizations of a minimally rigid graph is a notoriously difficult problem. Towards this goal, for graphs that are minimally rigid in the plane, we take advantage of a recently published algorithm, which is the fastest available method, although its complexity is still exponential. Combining computational results with the theory of constructing new rigid graphs by gluing, in [4] we give a new lower bound on the maximal possible number of (complex) realizations for graphs with a given number of vertices. We extend these ideas to rigid graphs in three dimensions and we derive similar lower bounds, by exploiting data from extensive Gröbner basis computations.

### *6.2.4. The Complexity of Subdivision for Diameter-Distance Tests*

In [1] we present a general framework for analyzing the complexity of subdivision-based algorithms whose tests are based on the sizes of regions and their distance to certain sets (often varieties) intrinsic to the problem under study. We call such tests diameter-distance tests. We illustrate that diameter-distance tests are common in the literature by proving that many interval arithmetic-based tests are, in fact, diameter-distance tests. For this class of algorithms, we provide both non-adaptive bounds for the complexity, based on separation bounds, as well as adaptive bounds, by applying the framework of continuous amortization. Using this structure, we provide the first complexity analysis for the algorithm by Plantinga and Vegeter for approximating real implicit curves and surfaces. We present both adaptive and non-adaptive a priori worst-case bounds on the complexity of this algorithm both in terms of the number of subregions constructed and in terms of the bit complexity for the construction. Finally, we construct families of hypersurfaces to prove that our bounds are tight.

### *6.2.5. Real root finding for equivariant semi-algebraic systems*

Let $R$ be a real closed field. In [19] we consider basic semi-algebraic sets defined by $n$-variate equations/inequalities of s symmetric polynomials and an equivariant family of polynomials, all of them of degree bounded by $2d < n$. Such a semi-algebraic set is invariant by the action of the symmetric group. We show that such a set is either empty or it contains a point with at most $2d-1$ distinct coordinates. Combining this geometric result with efficient algorithms for real root finding (based on the critical point method), one can decide the emptiness of basic semi-algebraic sets defined by $s$ polynomials of degree $d$ in time $(sn)^{O(d)}$. This improves the state-of-the-art which is exponential in $n$. When the variables $x_1, ..., x_n$ are quantified and the coefficients of the input system depend on parameters $y_1, ..., y_t$, one also demonstrates that the corresponding one-block quantifier elimination problem can be solved in time $(sn)^{O(dt)}$.

### *6.2.6. Exact algorithms for semidefinite programs with degenerate feasible set*

Let $A_0, ..., A_n n$ be $m \times m$ symmetric matrices with entries in $Q$, and let $A(x)$ be the linear pencil $A_0 + x_1A_1 + \cdots + x_nA_n$, where $x = (x1, ..., xn)$ are unknowns. The linear matrix inequality (LMI) $A(x) \succeq 0$ defines the subset of $R^n$, called spectrahedron, containing all points $x$ such that $A(x)$ has nonnegative eigenvalues. The minimization of linear functions over spectrahedra is called semidefinite programming (SDP). Such problems appear frequently in control theory and real algebra, especially in the context of nonnegativity certificates for multivariate polynomials based on sums of squares. Numerical software for solving SDP are mostly based on the interior point method, assuming some non-degeneracy properties such as the existence of interior points in the admissible set. In [21], we design an exact algorithm based on symbolic homotopy for solving semidefinite programs without assumptions on the feasible set, and we analyze its complexity. Because of the exactness of the output, it cannot compete with numerical routines in practice but we prove that solving such problems can be done in polynomial time if either $n$ or $m$ is fixed.

### *6.2.7. A lower bound on the positive semidefinite rank of convex bodies*

The positive semidefinite rank of a convex body $C$ is the size of its smallest positive semidef-inite formulation. In [3] we show that the positive semidefinite rank of any convex body $C$ is at least $\sqrt{\log d}$ where $d$ is the smallest degree of a polynomial that vanishes on the boundary of the polar of $C$. This improves on the existing bound which relies on results from quantifier elimination. Our proof relies on the Bézout bound applied to the Karush-Kuhn-Tucker conditions of optimality. We discuss the connection with the algebraic degree of

semidefinite programming and show that the bound is tight (up to constant factor) for random spectrahedra of suitable dimension.

### 6.2.8. On the complexity of computing real radicals of polynomial systems

Let $f = (f_1, ..., f_s)$ be a sequence of polynomials in $Q[X_1, ..., X_n]$ of maximal degree $D$ and $V \subset C^n$ be the algebraic set defined by $f$ and $r$ be its dimension. The real radical $\sqrt[re]{\langle f \rangle}$ associated to $f$ is the largest ideal which defines the real trace of $V$. In [20] when $V$ is smooth, we show that $\sqrt[re]{\langle f \rangle}$ has a finite set of generators with degrees bounded by V. Moreover, we present a probabilistic algorithm of complexity $(snDn)^{O(1)}$ to compute the minimal primes of $\sqrt[re]{\langle f \rangle}$ . When $V$ is not smooth, we give a probabilistic algorithm of complexity $s^{O(1)}(nD)^{O(nr2^r)}$ to compute rational parametrizations for all irreducible components of the real algebraic set $V \cap R^n$. Experiments are given to show the efficiency of our approaches.

### 6.2.9. Algorithms for Weighted Sums of Squares Decomposition of Non-negative Univariate Polynomials

It is well-known that every non-negative univariate real polynomial can be written as the sum of two polynomial squares with real coefficients. When one allows a weighted sum of finitely many squares instead of a sum of two squares, then one can choose all coefficients in the representation to lie in the field generated by the coefficients of the polynomial. In particular, this allows an effective treatment of polynomials with rational coefficients. In [9], we describe, analyze and compare both from the theoretical and practical points of view, two algorithms computing such a weighted sums of squares decomposition for univariate polynomials with rational coefficients. The first algorithm, due to the third author relies on real root isolation, quadratic approximations of positive polynomials and square-free decomposition but its complexity was not analyzed. We provide bit complexity estimates, both on the runtime and the output size of this algorithm. They are exponential in the degree of the input univariate polynomial and linear in the maximum bitsize of its complexity. This analysis is obtained using quantifier elimination and root isolation bounds. The second algorithm, due to Chevillard, Harrison, Joldes and Lauter, relies on complex root isolation and square-free decomposition and has been introduced for certifying positiveness of poly-nomials in the context of computer arithmetics. Again, its complexity was not analyzed. We provide bit complexity estimates, both on the runtime and the output size of this algorithm, which are polynomial in the degree of the input polynomial and linear in the maximum bitsize of its complexity. This analysis is obtained using Vieta's formula and root isolation bounds. Finally, we report on our implementations of both algorithms and compare them in practice on several application benchmarks. While the second algorithm is, as expected from the complexity result, more efficient on most of examples, we exhibit families of non-negative polynomials for which the first algorithm is better.

### 6.2.10. On Exact Polya and Putinar's Representations

We consider the problem of finding exact sums of squares (SOS) decompositions for certain classes of non-negative multivariate polynomials, relying on semidefinite programming (SDP) solvers. In [18] we start by providing a hybrid numeric-symbolic algorithm computing exact rational SOS decompositions for polynomials lying in the interior of the SOS cone. It computes an approximate SOS decomposition for a perturbation of the input polynomial with an arbitrary-precision SDP solver. An exact SOS decomposition is obtained thanks to the perturbation terms. We prove that bit complexity estimates on output size and runtime are both polynomial in the degree of the input polynomial and simply exponential in the number of variables. Next, we apply this algorithm to compute exact Polya and Putinar's representations respectively for positive definite forms and positive polynomials over basic compact semi-algebraic sets. We also compare the implementation of our algorithms with existing methods in computer algebra including cylindrical algebraic decomposition and critical point method.

## 6.3. Solving Systems in Finite Fields, Applications in Cryptology and Algebraic Number Theory.

### 6.3.1. Linear Repairing Codes and Side-Channel Attacks

To strengthen the resistance of countermeasures based on secret sharing, several works have suggested to use the scheme introduced by Shamir in 1978, which proposes to use the evaluation of a random d-degree polynomial into $nd + 1$ public points to share the sensitive data. Applying the same principles used against the classical Boolean sharing, all these works have assumed that the most efficient attack strategy was to exploit the minimum number of shares required to rebuild the sensitive value; which is $d + 1$ if the reconstruction is made with Lagrange's interpolation. In [2], we highlight first an important difference between Boolean and Shamir's sharings which implies that, for some signal-to-noise ratio, it is more advantageous for the adversary to observe strictly more than d + 1 shares. We argue that this difference is related to the existence of so-called exact linear repairing codes, which themselves come with reconstruction formulae that need (much) less information (counted in bits) than Lagrange's interpolation. In particular, this result implies that, contrary to what was believed, the choice of the public points in Shamir's sharing has an impact on the countermeasure strength. As another contribution, we exhibit a positive impact of the existence of linear exact repairing schemes; we indeed propose to use them to improve the state-of-the-art multiplication algorithms dedicated to Shamir's sharing. We argue that the improvement can be effective when the multiplication operation in the base field is at least two times smaller than in its sub-fields.

### *6.3.2. On the Use of Independent Component Analysis to Denoise Side-Channel Measurements*

Independent Component Analysis (ICA) is a powerful technique for blind source separation. It has been successfully applied to signal processing problems, such as feature extraction and noise reduction , in many different areas including medical signal processing and telecommunication. In [17], we propose a framework to apply ICA to denoise side-channel measurements and hence to reduce the complexity of key recovery attacks. Based on several case studies, we afterwards demonstrate the overwhelming advantages of ICA with respect to the commonly used preprocessing techniques such as the singular spectrum analysis. Mainly, we target a software masked implementation of an AES and a hardware unprotected one. Our results show a significant Signal-to-Noise Ratio (SNR) gain which translates into a gain in the number of traces needed for a successful side-channel attack. This states the ICA as an important new tool for the security assessment of cryptographic implementations.

# 7. Bilateral Contracts and Grants with Industry

## 7.1. Bilateral Grants with Industry

Until the mid 2000's, multivariate cryptography was developing very rapidly, producing many interesting and versatile public-key schemes. However, many of them were soon successfully cryptanalysed (a lot have been done in this group). As a consequence, the confidence in multivariate cryptography cryptosystems declined. It seems that there have emerged new important reasons for renewal of the interest in a new generation of multivariate schemes. In the past two years, the algorithms for solving the Discrete Logarithm Problem over small characteristic fields underwent an extraordinary development. This clearly illustrates the risk to not consider alternatives to classical assumptions based on number theory. In parallel, two of the most important standardization bodies in the world, NIST and ETSI have recently started initiatives for developing cryptographic standards not based on number theory, with a particular focus on primitives resistant to quantum algorithms. An objective here is then to focus on the design of multivariate schemes.

The team is involved in the industrial transfer of post-quantum cryptography. The maturation project, called HFEBOOST, is supervised by SATT-LUTECH.

SATT-LUTECH specializes in the processing and transfer of technologies from research laboratories of its shareholders: Inria, CNRS, University of Technology of Compiègne National Museum of Natural History, Institute Curie, Université Panthéon-Assas, Paris Sorbonne University and National School of Industrial Creation).

The team has recently developed, in partnership with a mobile application development company (WASSA), an Android app for smartphones (Samsung S5 type) that uses multivariate cryptography. The application has been tested mid-November in a series of experiments supervised by DGA and French Ministry of Defense. The experiment gathered a total of hundred participants from various operational units. This is a first milestone in the maturation project whose goal is to create a start-up.

## 7.2. Public Contracts

CEA LETI / DSYS / CESTI

In smart card domain, the emanations of a component during a cryptographic computation may compromise the information that is directly or not linked to the secret keys. The most part of the side channel attacks are based on statistical tools that exploit relations between the handled data and the signals. However these methods do not take advantage of all the signal information. The goal is to study the existing algorithms in pattern and speech recognition and to apply them to signals related to cryptographic computations. The objective will be to improve the attacks efficiency and resolve more complex problems.

- CIFRE Contract with ST Micro electronics that funds the PhD thesis of Simon Landry on "Threshold Implementations Against Side Channel Analysis". Supervisor Emmanuel Prouff.

# 8. Partnerships and Cooperations

## 8.1. Regional Initiatives

- **French Ministry of Armies**

  POLSYS has a collaboration with the French Ministry of Armies.

- **Grant GAMMA** (funded by PGMO).

  GLOBAL ALGEBRAIC SHOOTING METHOD IN OPTIMAL CONTROL AND APPLICATIONS

  Optimal control consists in steering a system from an initial configuration to a final one, while minimizing some given cost criterion. One of the current main challenges is to develop innovative methods for computing global solutions. This is crucial for applications where validating the global control laws is a crucial but a highly time consuming and expensive phase. GAMMA focuses on the wide range of optimal control problems having an algebraic structure, involving for instance polynomial or semi-algebraic dynamics and costs, or switches between polynomial models. In this case, GAMMA aims at designing methods relying on algebraic computations to the mainstream shooting method in order to yield optimal solutions that purely numerical techniques cannot provide.

## 8.2. National Initiatives

### *8.2.1. ANR*

- **ANR Jeunes Chercheurs GALOP (Games through the lens of ALgebra and OPptimization)**

  `Duration:` 2018–2022

  GALOP [0] is a Young Researchers (JCJC) project with the purpose of extending the limits of the state-of-the-art algebraic tools in computer science, especially in stochastic games. It brings original and innovative algebraic tools, based on symbolic-numeric computing, that exploit the geometry and the structure and complement the state-of-the-art. We support our theoretical tools with a highly efficient open-source software for solving polynomials. Using our algebraic tools we study the geometry of the central curve of (semi-definite) optimization problems. The algebraic tools and our results from the geometry of optimization pave the way to introduce algorithms and precise bounds for stochastic games.

[0] https://project.inria.fr/galop/

Participants: E. Tsigaridas [contact], F. Johansson, H. Gimbert, J.-C. Faugère, M. Safey El Din.

### *8.2.2. Programme d'investissements d'avenir (PIA)*

- **PIA grant RISQ: Regroupement of the Security Industry for Quantum-Safe security (2017-2020).** The goal of the RISQ project is to prepare the security industry to the upcoming shift of classical cryptography to quantum-safe cryptography. (J.-C. Faugère [contact], and L. Perret).

  The RISQ [0] project is certainly the biggest industrial project ever organized in quantum-safe cryptography. RISQ is one of few projects accepted in the call Grands DΓ©fis du NumΓ©rique which is managed by BPI France, and will be funded thanks to the so-called Plan d'Investissements d'Avenir.

  The RISQ project is a natural continuation of POLSYS commitment to the industrial transfert of quantum-safe cryptography. RISQ is a large scale version of the HFEBoost project; which demonstrated the potential of quantum-safe cryptography.

  POLSYS actively participated to shape the RISQ project. POLSYS is now a member of the strategic board of RISQ, and is leading the task of designing and analyzing quantum-safe algorithms. In particular, a first milestone of this task was to prepare submissions to NIST's quantum-safe standardisation process.
- **ANR SESAME (Singularités Et Stabilité des AsservisseMEnts référencés capteurs)**

  Duration: 2018–2022

  Participants: J.-C. Faugère, M. Safey El Din.

## 8.3. European Initiatives

### *8.3.1. FP7 & H2020 Projects*

- **Innovative Training Network POEMA (Polynomial Optimization, Efficiency through Moments and Algebra)**

  Duration: 2019-2022.

  POEMA is a Marie Skłodowska-Curie Innovative Training Network (2019-2022).

  Its goal is to train scientists at the interplay of algebra, geometry and computer science for polynomial optimization problems and to foster scientific and technological advances, stimulating interdisciplinary and intersectoriality knowledge exchange between algebraists, geometers, computer scientists and industrial actors facing real-life optimization problems.

  Participants: J. Berthomieu, J.-C. Faugère, M. Safey El Din [contact], E. Tsigaridas.

### *8.3.2. Collaborations in European Programs, Except FP7 & H2020*

Program: COST

Project acronym: CryptoAction

Project title: Cryptography for Secure Digital Interaction

Duration: Apr. 2014 - Apr. 2018

Coordinator: Claudio ORLANDI

Abstract: As increasing amounts of sensitive data are exchanged and processed every day on the Internet, the need for security is paramount. Cryptography is the fundamental tool for securing digital interactions, and allows much more than secure communication: recent breakthroughs in cryptography enable the protection - at least from a theoretical point of view - of any interactive data processing task. This includes electronic voting, outsourcing of storage and computation, e-payments, electronic auctions, etc. However, as cryptography advances and becomes more complex,

[0] http://risq.fr/

single research groups become specialized and lose contact with "the big picture". Fragmentation in this field can be dangerous, as a chain is only as strong as its weakest link. To ensure that the ideas produced in Europe's many excellent research groups will have a practical impact, coordination among national efforts and different skills is needed. The aim of this COST Action is to stimulate interaction between the different national efforts in order to develop new cryptographic solutions and to evaluate the security of deployed algorithms with applications to the secure digital interactions between citizens, companies and governments. The Action will foster a network of European research centers thus promoting movement of ideas and people between partners.

Program: COST

Project acronym: CRYPTACUS

Project title: Cryptanalysis of ubiquitous computing systems

Duration: Dec. 2014 - Dec. 2018

Coordinator: Gildas AVOINE

Abstract: Recent technological advances in hardware and software have irrevocably affected the classical picture of computing systems. Today, these no longer consist only of connected servers, but involve a wide range of pervasive and embedded devices, leading to the concept of "ubiquitous computing systems". The objective of the Action is to improve and adapt the existent cryptanalysis methodologies and tools to the ubiquitous computing framework. Cryptanalysis, which is the assessment of theoretical and practical cryptographic mechanisms designed to ensure security and privacy, will be implemented along four axes: cryptographic models, cryptanalysis of building blocks, hardware and software security engineering, and security assessment of real-world systems. Researchers have only recently started to focus on the security of ubiquitous computing systems. Despite the critical flaws found, the required highly-specialized skills and the isolation of the involved disciplines are a true barrier for identifying additional issues. The Action will establish a network of complementary skills, so that expertise in cryptography, information security, privacy, and embedded systems can be put to work together. The outcome will directly help industry stakeholders and regulatory bodies to increase security and privacy in ubiquitous computing systems, in order to eventually make citizens better protected in their everyday life.

## 8.4. International Research Visitors

### *8.4.1. Visits of International Scientists*

*8.4.1.1. Internships*

Reine Abi Rached

Date: Apr. 2018 - Aug. 2018

Institution: Université de Versailles –St-Quentin-en-Yvelines

Supervisor: Jean-Charles Faugère, Jérémy Berthomieu

Hadrien Brochet

Date: Jun. 2018 - Aug. 2018

Institution: ENS Lyon

Supervisor: Elias Tsigaridas

Phuoc Le

Date: Apr. 2018 - Aug. 2018

Institution: Université de Versailles –St-Quentin-en-Yvelines

Supervisor: Jean-Charles Faugère, Mohab Safey El Din

### *8.4.2. Visits to International Teams*

*8.4.2.1. Research Stays Abroad*

Elias Tsigaridas was a visiting research scientist at the ICERM institute (Brown University) during the special semester on "Nonlinear Algebra" (Sep – Nov 2018).

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Organisation*

*9.1.1.1. General Chair, Scientific Chair*

Dongming Wang was the General Chair of International Conference on Automated Deduction in Geometry (ADG 2018) (Nanning, China, September 11-14, 2018).

Dongming Wang was the General co-Chair of the 44th International Symposium on Symbolic and Algebraic Computation (ISSAC 2019) , Beijing, China, July 15-18, 2019), and the 13th International Conference on Artificial Intelligence and Symbolic Computation (AISC 2018) (Suzhou, China, September 16-19, 2018).

### *9.1.2. Scientific Events Selection*

*9.1.2.1. Member of the Conference Program Committees*

Elias Tsigaridas was a member of the program committees of the 20th International Workshop on Computer Algebra in Scientific Computing (CASC) 2018.

Mohab Safey El Din was member of the program committee of the 43rd International Symposium on Symbolic and Algebraic Computation (ISSAC) 2018.

Emmanuel Prouff was a member of the programm committee of the Conference on Cryptographic Hardware and Embedded Systems 2018 (CHES), Smart Card Research and Advanced Application Conference (CARDIS) 2018, and Constructive Side-Channel Analysis and Secure Design (COSADE) 2018.

Dongming Wang was a member of the program committee of 13th International Conference on Artificial Intelligence and Symbolic Computation (AISC 2018) (Suzhou, China, September 16-19, 2018) and the 4th International Conference on Numerical and Symbolic Computation (SYMCOMP 2019) (Porto, Portugal, April 11-12, 2019).

*9.1.2.2. Reviewer*

Mohab Safey El Din was reviewer of the M. Skomra's Phd (CMAP, École polytechnique).

### *9.1.3. Journal*

*9.1.3.1. Member of the Editorial Boards*

Mohab Safey El Din is member of the editorial board of the Journal of Symbolic Computation.

Mohab Safey El Din (with Chee Yap, Courant Inst. NYU) is guest editor of the Journal of Symbolic Computation Special Issue on the 2017 International Symposium on Symbolic and Algebraic Computation.

Dongming Wang is a member of the editorial board of

- Journal of Symbolic Computation (published by Academic Press/Elsevier, London),
- Frontiers of Computer Science (published by Higher Education Press, Beijing and Springer, Berlin),
- Texts and Monographs in Symbolic Computation (published by Springer, Wien New York).

Dongming Wang is a member of the Advisory Board for the journal SCIENCE CHINA Information Sciences (published by Science China Press, Beijing and Springer, Berlin).

Dongming Wang is the Editor-in-Chief for the journal Mathematics in Computer Science (published by Birkhäuser/Springer, Basel).

### *9.1.4. Invited Talks*

Elias Tsigaridas was invited speaker at

- IBM T.J. Watson Research Center, (*Invited talk*) 28 Nov 2018.
- *Applied Algebra Day*. MIT, 17 Nov 2018.
- ICERM, University of Brown, *Main seminar*, Nov, 2018.

Mohab Safey El Din was invited speaker at

- Key Lab on Math. Mechanization, Chinese Academy of Sciences, *Invited talk*.
- Dep. of Math. of Univ. of Tromso, *Invited talk*.
- ICERM, Semester Prog. on Non-linear Algebra, Workshop on Real algebraic geometry and optimization, *Plenary talk*.

Emmanuel Prouff was an invited speaker at

- PANDA 2018 Conference (China) and talked on "Deep Learning for Embedded Security Evaluation".
- COSADE 2018 Conference (Singapur) and talked on "Deep Learning for Embedded Security Evaluation".

### *9.1.5. Scientific Expertise*

Mohab Safey El Din is Chargé de Mission for Computer Science at Sorbonne Univ. (Faculté des Sciences et Ingéniérie).

## 9.2. Teaching - Supervision - Juries

### *9.2.1. Teaching*

Jérémy Berthomieu had the following teaching activities:

Master : Computation Modeling, 38 hours, M1, Sorbonne Université, France.

Master : In charge of Basics of Algebraic Algorithms, 74 hours, M1, Sorbonne Université & Polytech' UPMC, France.

Master : Projects supervision, 6 hours, M1, Sorbonne Université, France.

Licence : Introduction to Algorithmics, 33 hours, L2, Sorbonne Université , France.

Licence : Projects supervision, 10 hours, L2, Sorbonne Université, France.

Licence : In charge of Basics of Programmation 2, 50 hours, L1, Sorbonne Université, France.

Mohab Safey El Din has the following teaching activities:

Master : Computation Modeling, 33 hours, M1, Sorbonne Université, France.

Master : Polynomial System Solving, 40 hours, M1, Sorbonne Université, France.

Master : In charge of the curriculum on Security, Reliability of Performance in Computing, 30 hours, M1, Sorbonne Université , France.

Master : Projects management, 20 hours, M1, Sorbonne Université, France.

Licence : Projects supervision, 10 hours, L2, Sorbonne Université, France.

### *9.2.2. Supervision*

PhD in progress : Matías Bender, Algorithms for Sparse Gröbner basis and applications, started in Dec. 2015, Jean-Charles Faugère and Elias Tsigaridas.

PhD in progress : Thi Xuan Vu, Faster algorithms for structured polynomial systems, started in Oct. 2017, Jean-Charles Faugère and Mohab Safey El Din.

PhD in progress : Phuoc Le, Real root classification and polar varieties, started in Oct. 2018, Jean-Charles Faugère and Mohab Safey El Din.

PhD in progress : Simon Landry, Threshold Implementations Against Side Channel Analysis, Emmanuel Prouff.
CIFRE/Contract with ST Micro electronics.

### *9.2.3. Juries*

Mohab Safey El Din was member of the PhD committees of M. Skomra (CMAP, École polytechnique) and T. Weisser (LAAS, CNRS).

# 10. Bibliography

## Publications of the year

### Articles in International Peer-Reviewed Journal

[1] M. BURR, S. GAO, E. TSIGARIDAS.*The Complexity of Subdivision for Diameter-Distance Tests*, in "Journal of Symbolic Computation", 2018, https://hal.inria.fr/hal-01953446

[2] H. CHABANNE, H. MAGHREBI, E. PROUFF.*Linear Repairing Codes and Side-Channel Attacks*, in "IACR Transactions on Cryptographic Hardware and Embedded Systems", February 2018, vol. 2018, n° 1, p. 118-141 [*DOI :* 10.13154/TCHES.V2018.I1.118-141], https://hal.archives-ouvertes.fr/hal-01973360

[3] H. FAWZI, M. SAFEY EL DIN.*A lower bound on the positive semidefinite rank of convex bodies*, in "SIAM Journal on Applied Algebra and Geometry", 2018, vol. 2, n° 1, p. 126-139 [*DOI :* 10.1137/17M1142570], https://hal.inria.fr/hal-01657849

[4] G. GRASEGGER, C. KOUTSCHAN, E. TSIGARIDAS.*Lower bounds on the number of realizations of rigid graphs*, in "Experimental Mathematics", 2018, p. 1-22, https://hal.inria.fr/hal-01711441

[5] J.-B. B. LASSERRE, V. MAGRON.*Optimal data fitting: a moment approach*, in "SIAM Journal on Optimization", November 2018, vol. 28, n° 4, p. 3127-3144, https://arxiv.org/abs/1802.03259 - 21 pages, 5 figures [*DOI :* 10.1137/18M1170108], https://hal.archives-ouvertes.fr/hal-01706850

[6] V. MAGRON.*Interval Enclosures of Upper Bounds of Roundoff Errors using Semidefinite Programming*, in "ACM Transactions on Mathematical Software", August 2018, vol. 44, n° 4, p. 41:1–41:18, https://arxiv.org/abs/1611.01318 - 18 pages, 2 tables, 1 figure [*DOI :* 10.1145/3206430], https://hal.archives-ouvertes.fr/hal-01956815

[7] V. MAGRON, A. ROCCA, T. DANG.*Certified Roundoff Error Bounds using Bernstein Expansions and Sparse Krivine-Stengle Representations*, in "IEEE Transactions on Computers", 2018, https://arxiv.org/abs/1802.04385 - 14 pages, 2 figures, 2 tables. Extension of the work in arXiv:1610.07038 [*DOI :* 10.1109/TC.2018.2851235], https://hal.archives-ouvertes.fr/hal-01956817

[8] V. MAGRON, M. SAFEY EL DIN.*RealCertify: a Maple package for certifying non-negativity*, in "ACM Communications in Computer Algebra", June 2018, vol. 52, n° 2, p. 34-37, https://arxiv.org/abs/1805.02201 - 4 pages, 2 tables [*DOI :* 10.1145/3282678.3282681], https://hal.archives-ouvertes.fr/hal-01956812

[9] V. MAGRON, M. SAFEY EL DIN, M. SCHWEIGHOFER.*Algorithms for Weighted Sums of Squares Decomposition of Non-negative Univariate Polynomials*, in "Journal of Symbolic Computation", 2018, https://hal.archives-ouvertes.fr/hal-01538729

[10] M. SAFEY EL DIN, É. SCHOST.*Bit complexity for multi-homogeneous polynomial system solving Application to polynomial minimization*, in "Journal of Symbolic Computation", 2018, vol. 87, p. 176-206, https://arxiv.org/abs/1605.07433 [*DOI :* 10.1016/J.JSC.2017.08.001], https://hal.inria.fr/hal-01319729

[11] A. STRZEBONSKI, E. TSIGARIDAS.*Univariate real root isolation in an extension field and applications*, in "Journal of Symbolic Computation", 2018, https://hal.inria.fr/hal-01248390

### International Conferences with Proceedings

[12] L. BARTHELEMY, D. KAHROBAEI, G. RENAULT, Z. ŠUNIĆ.*Quadratic time algorithm for inversion of binary permutation polynomials*, in "ICMS 2018 - International Congress on Mathematical Software", South Bend, IN, United States, J. H. DAVENPO, M. KAUERS, G. LABAH, J. URBA (editors), Lecture Notes in Computer Science, Springer, July 2018, vol. 10931, p. 19-27 [*DOI :* 10.1007/978-3-319-96418-8_3], https://hal.archives-ouvertes.fr/hal-01981320

[13] E. BARTZOS, I. EMIRIS, J. LEGERSKÝ, E. TSIGARIDAS.*On the maximal number of real embeddings of spatial minimally rigid graphs*, in "ISSAC '18 International Symposium on Symbolic and Algebraic Computation", New York, United States, C. ARRECHE (editor), ACM, July 2018, p. 55-62 [*DOI :* 10.1145/3208976.3208994], https://hal.archives-ouvertes.fr/hal-01710518

[14] M. R. BENDER, J.-C. FAUGÈRE, A. MANTZAFLARIS, E. TSIGARIDAS.*Bilinear systems with two supports: Koszul resultant matrices, eigenvalues, and eigenvectors*, in "ISSAC 2018 - 43rd International Symposium on Symbolic and Algebraic Computation", New York, United States, July 2018, https://arxiv.org/abs/1805.05060 [*DOI :* 10.1145/3208976.3209011], https://hal.inria.fr/hal-01787549

[15] M. R. BENDER, J.-C. FAUGÈRE, E. TSIGARIDAS.*Towards Mixed Gröbner Basis Algorithms: the Multihomogeneous and Sparse Case*, in "ISSAC 2018 - 43rd International Symposium on Symbolic and Algebraic Computation", New York, United States, July 2018, https://arxiv.org/abs/1805.03577 [*DOI :* 10.1145/3208976.3209018], https://hal.inria.fr/hal-01787423

[16] J. BERTHOMIEU, J.-C. FAUGÈRE.*A Polynomial-Division-Based Algorithm for Computing Linear Recurrence Relations*, in "ISSAC 2018 - 43rd International Symposium on Symbolic and Algebraic Computation", New York, United States, July 2018 [*DOI :* 10.1145/3208976.3209017], https://hal.inria.fr/hal-01784369

[17] H. MAGHREBI, E. PROUFF.*On the Use of Independent Component Analysis to Denoise Side-Channel Measurements*, in "COSADE 2018 - 9th International Workshop on Constructive Side-Channel Analysis and Secure Design", Singapore, Singapore, Lecture Notes in Computer Science, Springer, April 2018, vol. 10815, p. 61-81 [*DOI :* 10.1007/978-3-319-89641-0_4], https://hal.archives-ouvertes.fr/hal-01973322

[18] V. MAGRON, M. SAFEY EL DIN.*On Exact Polya and Putinar's Representations*, in "ISSAC '18 International Symposium on Symbolic and Algebraic Computation", New-York, United States, ACM, July 2018, p. 279-286, https://arxiv.org/abs/1802.10339 - 19 pages, 4 algorithms, 3 tables [*DOI :* 10.1145/3208976.3208986], https://hal.archives-ouvertes.fr/hal-01720612

[19] C. RIENER, M. SAFEY EL DIN.*Real root finding for equivariant semi-algebraic systems*, in "ISSAC 20018 - 43rd International Symposium on Symbolic and Algebraic Computation", New-York, United States, July 2018, https://arxiv.org/abs/1806.08121 , https://hal.inria.fr/hal-01819106

[20] M. SAFEY EL DIN, Z.-H. YANG, L. ZHI.*On the complexity of computing real radicals of polynomial systems*, in "ISSAC '18 - The 2018 ACM on International Symposium on Symbolic and Algebraic Computation", New-York, United States, ACM, July 2018, p. 351-358 [*DOI :* 10.1145/3208976.3209002], https://hal.inria.fr/hal-01956596

### Conferences without Proceedings

[21] D. HENRION, S. NALDI, M. SAFEY EL DIN.*Exact algorithms for semidefinite programs with degenerate feasible set*, in "ISSAC 2018 - 43rd International Symposium on Symbolic and Algebraic Computation", New York City, United States, July 2018, 17, https://arxiv.org/abs/1802.02834 , https://hal.archives-ouvertes.fr/hal-01705590

## Other Publications

[22] J. G. ALCÁZAR, J. CARAVANTES, G. M. DIAZ-TOCA, E. TSIGARIDAS.*Computing the topology of a planar or space hyperelliptic curve*, January 2019, working paper or preprint, https://hal.inria.fr/hal-01968776

[23] M. R. BENDER, J.-C. FAUGÈRE, L. PERRET, E. TSIGARIDAS.*A nearly optimal algorithm to decompose binary forms*, June 2018, https://arxiv.org/abs/1810.12588 - In submission, https://hal.inria.fr/hal-01907777

[24] J. BERTHOMIEU, J.-C. FAUGÈRE.*A Polynomial-Division-Based Algorithm for Computing Linear Recurrence Relations*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01935229

[25] J. BERTHOMIEU, J.-C. FAUGÈRE.*In-depth comparison of the Berlekamp–Massey–Sakata and the Scalar-FGLM algorithms: the adaptive variants*, June 2018, https://arxiv.org/abs/1806.00978 - working paper or preprint, https://hal.inria.fr/hal-01805478

[26] J. D. HAUENSTEIN, M. SAFEY EL DIN, É. SCHOST, T. X. VU.*Solving determinantal systems using homotopy techniques*, February 2018, https://arxiv.org/abs/1802.10409 - working paper or preprint, https://hal.inria.fr/hal-01719170

[27] V. MAGRON, M. FORETS, D. HENRION.*Semidefinite Approximations of Invariant Measures for Polynomial Systems*, July 2018, https://arxiv.org/abs/1807.00754 - 28 pages, 14 figures, https://hal.archives-ouvertes.fr/hal-01828443

[28] V. MAGRON, M. SAFEY EL DIN.*On Exact Polya, Hilbert-Artin and Putinar's Representations*, November 2018, https://arxiv.org/abs/1811.10062 - 29 pages, 4 tables, extended version of the paper from ISSAC'18 conference (available at arXiv::1802.10339), https://hal.archives-ouvertes.fr/hal-01935727

[29] F. MORAIN, G. RENAULT, B. SMITH.*Deterministic factoring with oracles*, February 2018, https://arxiv.org/abs/1802.08444 - working paper or preprint, https://hal.inria.fr/hal-01715832

# Project-Team PROSECCO

## Programming securely with cryptography

RESEARCH CENTER
**Paris**

THEME
**Security and Confidentiality**

# Table of contents

# Project-Team PROSECCO

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

**Keywords:**

**Computer Science and Digital Science:**

A1.1. - Architectures
A1.1.8. - Security of architectures
A1.2. - Networks
A1.2.8. - Network security
A1.3. - Distributed Systems
A2. - Software
A2.1. - Programming Languages
A2.1.1. - Semantics of programming languages
A2.1.4. - Functional programming
A2.1.7. - Distributed programming
A2.1.11. - Proof languages
A2.2. - Compilation
A2.2.1. - Static analysis
A2.2.5. - Run-time systems
A2.4. - Formal method for verification, reliability, certification
A2.4.2. - Model-checking
A2.4.3. - Proofs
A2.5. - Software engineering
A4. - Security and privacy
A4.3. - Cryptography
A4.3.3. - Cryptographic protocols
A4.5. - Formal methods for security
A4.6. - Authentication
A4.8. - Privacy-enhancing technologies

**Other Research Topics and Application Domains:**

B6. - IT and telecom
B6.1. - Software industry
B6.1.1. - Software engineering
B6.3. - Network functions
B6.3.1. - Web
B6.3.2. - Network protocols
B6.4. - Internet of things
B9. - Society and Knowledge
B9.10. - Privacy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Karthikeyan Bhargavan [Team leader, Inria, Senior Researcher, HDR]
Amal Ahmed [Inria, Advanced Research Position, until Jul 2018]
Bruno Blanchet [Inria, Senior Researcher, HDR]
Harry Halpin [Inria, Starting Research Position]
Catalin Hritcu [Inria, Researcher]
Prasad Naldurg [Inria, Advanced Research Position, from May 2018]
Éric Tanter [Inria, Advanced Research Position, from Jul 2018]

**External Collaborators**

David Baelde [Ecole Normale Supérieure Cachan, until Aug 2018]
Théo Laurent [Ecole Normale Supérieure Paris, until Jun 2018]
Jean-Karim Zinzindohoué [Ministère de l'Intérieur, from Jul 2018]
Victor Dumitrescu [Nomadic Labs, from Dec 2018]

**Technical Staff**

Iness Ben Guirat [Inria, from Aug 2018]
Victor Dumitrescu [Inria, until Nov 2018]
Florian Groult [Inria, from Nov 2018]
Théo Laurent [Inria, from Jul 2018]
Denis Merigoux [Inria, from Feb 2018 until Jun 2018]
Marc Sylvestre [Inria, until Sep 2018]

**PhD Students**

Benjamin Beurdouche [Inria]
Nadim Kobeissi [Inria, until Jul 2018]
Natalia Kulatova [Inria]
Kenji Maillard [Ecole Normale Supérieure Paris]
Denis Merigoux [Inria, from Nov 2018]
Marina Polubelova [Inria]
Jean-Karim Zinzindohoué [Ministère de l'Intérieur, until Jun 2018]
Carmine Abate [Inria, from Jun 2018]
Benjamin Lipp [Inria, from Jun 2018]
Jérémy Thibault [Inria, from Aug 2018]

**Post-Doctoral Fellows**

Danel Ahman [Inria, until Sep 2018]
Roberto Blanco Martinez [Inria]

**Visiting Scientist**

Aaron Weiss [Northeastern University, until Jul 2018]

**Administrative Assistants**

Anna Bednarik [Inria, until Oct 2018]
Mathieu Mourey [Inria]

# 2. Overall Objectives

## 2.1. Programming securely with cryptography

In recent years, an increasing amount of sensitive data is being generated, manipulated, and accessed online, from bank accounts to health records. Both national security and individual privacy have come to rely on the security of web-based software applications. But even a single design flaw or implementation bug in an application may be exploited by a malicious criminal to steal, modify, or forge the private records of innocent users. Such *attacks* are becoming increasingly common and now affect millions of users every year.

The risks of deploying insecure software are too great to tolerate anything less than mathematical proof, but applications have become too large for security experts to examine by hand, and automated verification tools do not scale. Today, there is not a single widely-used web application for which we can give a proof of security, even against a small class of attacks. In fact, design and implementation flaws are still found in widely-distributed and thoroughly-vetted security libraries designed and implemented by experts.

Software security is in crisis. A focused research effort is needed if security programming and analysis techniques are to keep up with the rapid development and deployment of security-critical distributed applications based on new cryptographic protocols and secure hardware devices. The goal of our team PROSECCO is to draw upon our expertise in cryptographic protocols and program verification to make decisive contributions in this direction.

Our vision is that, over its lifetime, PROSECCO will contribute to making the use of formal techniques when programming with cryptography as natural as the use of a software debugger. To this end, our long-term goals are to design and implement programming language abstractions, cryptographic models, verification tools, and verified security libraries that developers can use to deploy provably secure distributed applications. Our target applications include cryptographic protocol implementations, hardware-based security APIs, smartphone- and browser-based web applications, and cloud-based web services. In particular, we aim to verify the full application: both the cryptographic core and the high-level application code. We aim to verify implementations, not just models. We aim to account for computational cryptography, not just its symbolic abstraction.

We identify five key focus areas for our research in the short- to medium term.

### *2.1.1. New programming languages for verified software*

Building realistic verified applications requires new programming languages that enable the systematic development of efficient software hand-in-hand with their proofs of correctness. Our current focus is on designing and implementing the programming language F*, in collaboration with Microsoft Research. F* (pronounced F star) is an ML-like functional programming language aimed at program verification. Its type system includes polymorphism, dependent types, monadic effects, refinement types, and a weakest precondition calculus. Together, these features allow expressing precise and compact specifications for programs, including functional correctness and security properties. The F* type-checker aims to prove that programs meet their specifications using a combination of SMT solving and manual proofs. Programs written in F* can be translated to efficient OCaml, F#, or C for execution. The main ongoing use case of F* is building a verified, drop-in replacement for the whole HTTPS stack in Project Everest (a larger collaboration with Microsoft Research). This includes verified implementations of TLS 1.2 and 1.3 and of the underlying cryptographic primitives.

### *2.1.2. Symbolic verification of cryptographic applications*

We aim to develop our own security verification tools for models and implementations of cryptographic protocols and security APIs using symbolic cryptography. Our starting point is the tools we have previously developed: the specialized cryptographic prover ProVerif, the reverse engineering and formal test tool Tookan, and the F* verification system. These tools are already used to verify industrial-strength cryptographic protocol implementations and commercial cryptographic hardware. We plan to extend and combine these approaches to capture more sophisticated attacks on applications consisiting of protocols, software, and hardware, as well as to prove symbolic security properties for such composite systems.

### *2.1.3. Computational verification of cryptographic applications*

We aim to develop our own cryptographic application verification tools that use the computational model of cryptography. The tools include the computational prover CryptoVerif, and the F* verification system. Working together, we plan to extend these tools to analyze, for the first time, cryptographic protocols, security APIs, and their implementations under fully precise cryptographic assumptions. We also plan to pursue links between symbolic and computational verification, such as computational soundness results that enable computational proofs by symbolic techniques.

### 2.1.4. Efficient formally secure compilers for tagged architectures

We aim to leverage emerging hardware capabilities for fine-grained protection to build the first, efficient secure compilation chains for realistic low-level programming languages (the C language, and Low* a safe subset of C embedded in F* for verification). These compilation chains will provide a secure semantics for all programs and will ensure that high-level abstractions cannot be violated even when interacting with untrusted low-level code. To achieve this level of security without sacrificing efficiency, our secure compilation chains target a tagged architecture, which associates a metadata tag to each word and efficiently propagates and checks tags according to software-defined rules.

### 2.1.5. Building provably secure web applications

We aim to develop analysis tools and verified libraries to help programmers build provably secure web applications. The tools will include static and dynamic verification tools for client- and server-side JavaScript web applications, their verified deployment within HTML5 websites and browser extensions, as well as type-preserving compilers from high-level applications written in F* to JavaScript. In addition, we plan to model new security APIs in browsers and smartphones and develop the first formal semantics for various HTML5 web standards. We plan to combine these tools and models to analyze the security of multi-party web applications, consisting of clients on browsers and smartphones, and servers in the cloud.

# 3. Research Program

## 3.1. Symbolic verification of cryptographic applications

Despite decades of experience, designing and implementing cryptographic applications remains dangerously error-prone, even for experts. This is partly because cryptographic security is an inherently hard problem, and partly because automated verification tools require carefully-crafted inputs and are not widely applicable. To take just the example of TLS, a widely-deployed and well-studied cryptographic protocol designed, implemented, and verified by security experts, the lack of a formal proof about all its details has regularly led to the discovery of major attacks (including several in PROSECCO) on both the protocol and its implementations, after many years of unsuspecting use.

As a result, the automated verification for cryptographic applications is an active area of research, with a wide variety of tools being employed for verifying different kinds of applications.

In previous work, we have developed the following three approaches:

- ProVerif: a symbolic prover for cryptographic protocol models
- Tookan: an attack-finder for PKCS#11 hardware security devices
- F*: a new language that enables the verification of cryptographic applications

### 3.1.1. Verifying cryptographic protocols with ProVerif

Given a model of a cryptographic protocol, the problem is to verify that an active attacker, possibly with access to some cryptographic keys but unable to guess other secrets, cannot thwart security goals such as authentication and secrecy [70]; it has motivated a serious research effort on the formal analysis of cryptographic protocols, starting with [65] and eventually leading to effective verification tools, such as our tool ProVerif.

To use ProVerif, one encodes a protocol model in a formal language, called the applied pi-calculus, and ProVerif abstracts it to a set of generalized Horn clauses. This abstraction is a small approximation: it just ignores the number of repetitions of each action, so ProVerif is still very precise, more precise than, say, tree automata-based techniques. The price to pay for this precision is that ProVerif does not always terminate; however, it terminates in most cases in practice, and it always terminates on the interesting class of *tagged protocols* [60]. ProVerif can handle a wide variety of cryptographic primitives, defined by rewrite rules or by some equations, and prove a wide variety of security properties: secrecy [58], [44], correspondences (including authentication) [59], and observational equivalences [57]. Observational equivalence means that an adversary cannot distinguish two processes (protocols); equivalences can be used to formalize a wide range of properties, but they are particularly difficult to prove. Even if the class of equivalences that ProVerif can prove is limited to equivalences between processes that differ only by the terms they contain, these equivalences are useful in practice and ProVerif has long been the only tool that proves equivalences for an unbounded number of sessions. (Maude-NPA in 2014 and Tamarin in 2015 adopted ProVerif's approach to proving equivalences.)

Using ProVerif, it is now possible to verify large parts of industrial-strength protocols, such as TLS [52], Signal [68], JFK [45], and Web Services Security [56], against powerful adversaries that can run an unlimited number of protocol sessions, for strong security properties expressed as correspondence queries or equivalence assertions. ProVerif is used by many teams at the international level, and has been used in more than 120 research papers (references available at http://proverif.inria.fr/proverif-users.html).

### *3.1.2. Verifying security APIs using Tookan*

Security application programming interfaces (APIs) are interfaces that provide access to functionality while also enforcing a security policy, so that even if a malicious program makes calls to the interface, certain security properties will continue to hold. They are used, for example, by cryptographic devices such as smartcards and Hardware Security Modules (HSMs) to manage keys and provide access to cryptographic functions whilst keeping the keys secure. Like security protocols, their design is security critical and very difficult to get right. Hence formal techniques have been adapted from security protocols to security APIs.

The most widely used standard for cryptographic APIs is RSA PKCS#11, ubiquitous in devices from smartcards to HSMs. A 2003 paper highlighted possible flaws in PKCS#11 [62], results which were extended by formal analysis work using a Dolev-Yao style model of the standard [63]. However at this point it was not clear to what extent these flaws affected real commercial devices, since the standard is underspecified and can be implemented in many different ways. The Tookan tool, developed by Steel in collaboration with Bortolozzo, Centenaro and Focardi, was designed to address this problem. Tookan can reverse engineer the particular configuration of PKCS#11 used by a device under test by sending a carefully designed series of PKCS#11 commands and observing the return codes. These codes are used to instantiate a Dolev-Yao model of the device's API. This model can then be searched using a security protocol model checking tool to find attacks. If an attack is found, Tookan converts the trace from the model checker into the sequence of PKCS#11 queries needed to make the attack and executes the commands directly on the device. Results obtained by Tookan are remarkable: of 18 commercially available PKCS#11 devices tested, 10 were found to be susceptible to at least one attack.

### *3.1.3. Verifying cryptographic applications using F**

Verifying the implementation of a protocol has traditionally been considered much harder than verifying its model. This is mainly because implementations have to consider real-world details of the protocol, such as message formats, that models typically ignore. This leads to a situation that a protocol may have been proved secure in theory, but its implementation may be buggy and insecure. However, with recent advances in both program verification and symbolic protocol verification tools, it has become possible to verify fully functional protocol implementations in the symbolic model. One approach is to extract a symbolic protocol model from an implementation and then verify the model, say, using ProVerif. This approach has been quite successful, yielding a verified implementation of TLS in F# [55]. However, the generated models are typically quite large and whole-program symbolic verification does not scale very well.

An alternate approach is to develop a verification method directly for implementation code, using well-known program verification techniques. Our current focus is on designing and implementing the programming language F* [73], [49], in collaboration with Microsoft Research. F* (pronounced F star) is an ML-like functional programming language aimed at program verification. Its type system includes polymorphism, dependent types, monadic effects, refinement types, and a weakest precondition calculus. Together, these features allow expressing precise and compact specifications for programs, including functional correctness and security properties. The F* type-checker aims to prove that programs meet their specifications using a combination of SMT solving and manual proofs. Programs written in F* can be translated to efficient OCaml, F#, or C for execution [71]. The main ongoing use case of F* is building a verified, drop-in replacement for the whole HTTPS stack in Project Everest [53] (a larger collaboration with Microsoft Research). This includes a verified implementation of TLS 1.2 and 1.3 [54].

## 3.2. Computational verification of cryptographic applications

Proofs done by cryptographers in the computational model are mostly manual. Our goal is to provide computer support to build or verify these proofs. In order to reach this goal, we have designed the automatic tool CryptoVerif, which generates proofs by sequences of games. We already applied it to important protocols such as TLS [52] and Signal [68] but more work is still needed in order to develop this approach, so that it is easier to apply to more protocols. We also design and implement techniques for proving implementations of protocols secure in the computational model. In particular, CryptoVerif can generate implementations from CryptoVerif specifications that have been proved secure [61]. We plan to continue working on this approach.

A different approach is to directly verify cryptographic applications in the computational model by typing. A recent work [66] shows how to use refinement typechecking in F7 to prove computational security for protocol implementations. In this method, henceforth referred to as computational F7, typechecking is used as the main step to justify a classic game-hopping proof of computational security. The correctness of this method is based on a probabilistic semantics of F# programs and crucially relies on uses of type abstraction and parametricity to establish strong security properties, such as indistinguishability.

In principle, the two approaches, typechecking and game-based proofs, are complementary. Understanding how to combine these approaches remains an open and active topic of research.

An alternative to direct computation proofs is to identify the cryptographic assumptions under which symbolic proofs, which are typically easier to derive automatically, can be mapped to computational proofs. This line of research is sometimes called computational soundness and the extent of its applicability to real-world cryptographic protocols is an active area of investigation.

## 3.3. F*: A Higher-Order Effectful Language for Program Verification

F* [73], [49] is a verification system for effectful programs developed collaboratively by Inria and Microsoft Research. It puts together the automation of an SMT-backed deductive verification tool with the expressive power of a proof assistant based on dependent types. After verification, F* programs can be extracted to efficient OCaml, F#, or C code [71]. This enables verifying the functional correctness and security of realistic applications. F*'s type system includes dependent types, monadic effects, refinement types, and a weakest precondition calculus. Together, these features allow expressing precise and compact specifications for programs, including functional correctness and security properties. The F* type-checker aims to prove that programs meet their specifications using a combination of SMT solving and interactive proofs. The main ongoing use case of F* is building a verified, drop-in replacement for the whole HTTPS stack in Project Everest. This includes verified implementations of TLS 1.2 and 1.3 [54] and of the underlying cryptographic primitives [74].

## 3.4. Efficient Formally Secure Compilers to a Tagged Architecture

Severe low-level vulnerabilities abound in today's computer systems, allowing cyber-attackers to remotely gain full control. This happens in big part because our programming languages, compilers, and architectures

were designed in an era of scarce hardware resources and too often trade off security for efficiency. The semantics of mainstream low-level languages like C is inherently insecure, and even for safer languages, establishing security with respect to a high-level semantics does not guarantee the absence of low-level attacks. Secure compilation using the coarse-grained protection mechanisms provided by mainstream hardware architectures would be too inefficient for most practical scenarios.

We aim to leverage emerging hardware capabilities for fine-grained protection to build the first, efficient secure compilation chains for realistic low-level programming languages (the C language, and Low* a safe subset of C embedded in F* for verification [71]). These compilation chains will provide a secure semantics for all programs and will ensure that high-level abstractions cannot be violated even when interacting with untrusted low-level code. To achieve this level of security without sacrificing efficiency, our secure compilation chains target a tagged architecture [50], which associates a metadata tag to each word and efficiently propagates and checks tags according to software-defined rules. We hope to experimentally evaluate and carefully optimize the efficiency of our secure compilation chains on realistic workloads and standard benchmark suites. We are also using property-based testing and formal verification to provide high confidence that our compilation chains are indeed secure. Formally, we are constructing machine-checked proofs of a new security criterion we call robustly safe compilation, which is defined as the preservation of safety properties even against an adversarial context [46], [47]. This strong criterion complements compiler correctness and ensures that no machine-code attacker can do more harm to securely compiled components than a component already could with respect to a secure source-level semantics.

## 3.5. Provably secure web applications

Web applications are fast becoming the dominant programming platform for new software, probably because they offer a quick and easy way for developers to deploy and sell their *app*s to a large number of customers. Third-party web-based apps for Facebook, Apple, and Google, already number in the hundreds of thousands and are likely to grow in number. Many of these applications store and manage private user data, such as health information, credit card data, and GPS locations. To protect this data, applications tend to use an ad hoc combination of cryptographic primitives and protocols. Since designing cryptographic applications is easy to get wrong even for experts, we believe this is an opportune moment to develop security libraries and verification techniques to help web application programmers.

As a typical example, consider commercial password managers, such as LastPass, RoboForm, and 1Password. They are implemented as browser-based web applications that, for a monthly fee, offer to store a user's passwords securely on the web and synchronize them across all of the user's computers and smartphones. The passwords are encrypted using a master password (known only to the user) and stored in the cloud. Hence, no-one except the user should ever be able to read her passwords. When the user visits a web page that has a login form, the password manager asks the user to decrypt her password for this website and automatically fills in the login form. Hence, the user no longer has to remember passwords (except her master password) and all her passwords are available on every computer she uses.

Password managers are available as browser extensions for mainstream browsers such as Firefox, Chrome, and Internet Explorer, and as downloadable apps for Android and Apple phones. So, seen as a distributed application, each password manager application consists of a web service (written in PHP or Java), some number of browser extensions (written in JavaScript), and some smartphone apps (written in Java or Objective C). Each of these components uses a different cryptographic library to encrypt and decrypt password data. How do we verify the correctness of all these components?

We propose three approaches. For client-side web applications and browser extensions written in JavaScript, we propose to build a static and dynamic program analysis framework to verify security invariants. To this end, we have developed two security-oriented type systems for JavaScript, Defensive JavaScript [64] [64] and TS* [72], and used them to guarantee security properties for a number of JavaScript applications. For Android smartphone apps and web services written in Java, we propose to develop annotated JML cryptography libraries that can be used with static analysis tools like ESC/Java to verify the security of application code. For clients and web services written in F# for the .NET platform, we propose to use F* to verify their correctness.

We also propose to translate verified F* web applications to JavaScript via a verified compiler that preserves the semantics of F* programs in JavaScript.

### 3.6. Design and Verification of next-generation protocols: identity, blockchains, and messaging

Building on our work on verifying and re-designing pre-existing protocols like TLS and Web Security in general, with the resources provided by the NEXTLEAP project, we are working on both designing and verifying new protocols in rapidly emerging areas like identity, blockchains, and secure messaging. These are all areas where existing protocols, such as the heavily used OAuth protocol, are in need of considerable re-design in order to maintain privacy and security properties. Other emerging areas, such as blockchains and secure messaging, can have modifications to existing pre-standard proposals or even a complete 'clean slate' design. As shown by Prosecco's work, newer standards, such as IETF OAuth, W3C Web Crypto, and W3C Web Authentication API, can have vulnerabilities fixed before standardization is complete and heavily deployed. We hope that the tools used by Prosecco can shape the design of new protocols even before they are shipped to standards bodies. We have seen considerable progress in identity with the UnlimitID design and with messaging via the IETF MLS effort, with new work on blockchain technology underway.

## 4. Application Domains

### 4.1. Cryptographic Protocol Libraries

Cryptographic protocols such as TLS, SSH, IPSec, and Kerberos are the trusted base on which the security of modern distributed systems is built. Our work enables the analysis and verification of such protocols, both in their design and implementation. Hence, for example, we build and verify models and reference implementations for well-known protocols such as TLS and SSH, as well as analyze their popular implementations such as OpenSSL.

### 4.2. Hardware-based security APIs

Cryptographic devices such as Hardware Security Modules (HSMs) and smartcards are used to protect long-terms secrets in tamper-proof hardware, so that even attackers who gain physical access to the device cannot obtain its secrets. These devices are used in a variety of scenarios ranging from bank servers to transportation cards (e.g. Navigo). Our work investigates the security of commercial cryptographic hardware and evaluates the APIs they seek to implement.

### 4.3. Web application security

Web applications use a variety of cryptographic techniques to securely store and exchange sensitive data for their users. For example, a website may serve pages over HTTPS, authenticate users with a single sign-on protocol such as OAuth, encrypt user files on the server-side using XML encryption, and deploy client-side cryptographic mechanisms using a JavaScript cryptographic library. The security of these applications depends on the public key infrastructure (X.509 certificates), web browsers' implementation of HTTPS and the same origin policy (SOP), the semantics of JavaScript, HTML5, and their various associated security standards, as well as the correctness of the specific web application code of interest. We build analysis tools to find bugs in all these artifacts and verification tools that can analyze commercial web applications and evaluate their security against sophisticated web-based attacks.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

- We published 20 papers at top-tier conferences and journals such as POPL (5), ICFP (2), PLDI (1), OOPSLA (1), ACM CCS (1), IEEE S&P (1), IEEE CSF (1), TOPLAS (1), and JCS (1).
- The HACL* verified cryptographic library developed in our group was integrated by Linux (WireGuard) and Tezos, and more verified crypto primitives were integrated in Mozilla Firefox.
- We organized a Dagstuhl Seminar on Secure Compilation (18201)
- Catalin Hritcu served as Program Chair for the Workshop on Principles of Secure Compilation at POPL'18

# 6. New Software and Platforms

## 6.1. Cryptosense Analyzer

SCIENTIFIC DESCRIPTION: Cryptosense Analyzer (formerly known as Tookan) is a security analysis tool for cryptographic devices such as smartcards, security tokens and Hardware Security Modules that support the most widely-used industry standard interface, RSA PKCS#11. Each device implements PKCS#11 in a slightly different way since the standard is quite open, but finding a subset of the standard that results in a secure device, i.e. one where cryptographic keys cannot be revealed in clear, is actually rather tricky. Cryptosense Analyzer analyses a device by first reverse engineering the exact implementation of PKCS#11 in use, then building a logical model of this implementation for a model checker, calling a model checker to search for attacks, and in the case where an attack is found, executing it directly on the device. It has been used to find at least a dozen previously unknown flaws in commercially available devices.
FUNCTIONAL DESCRIPTION: Cryptosense Analyzer (formerly known as Tookan) is a security analysis tool for cryptographic devices such as smartcards,

- Participants: Graham Steel and Romain Bardou
- Contact: Graham Steel
- URL: https://cryptosense.com/

## 6.2. CryptoVerif

*Cryptographic protocol verifier in the computational model*
KEYWORDS: Security - Verification - Cryptographic protocol
FUNCTIONAL DESCRIPTION: CryptoVerif is an automatic protocol prover sound in the computational model. In this model, messages are bitstrings and the adversary is a polynomial-time probabilistic Turing machine. CryptoVerif can prove secrecy and correspondences, which include in particular authentication. It provides a generic mechanism for specifying the security assumptions on cryptographic primitives, which can handle in particular symmetric encryption, message authentication codes, public-key encryption, signatures, hash functions, and Diffie-Hellman key agreements. It also provides an explicit formula that gives the probability of breaking the protocol as a function of the probability of breaking each primitives, this is the exact security framework.
NEWS OF THE YEAR: Bruno Blanchet modified ProVerif and CryptoVerif to improve the compatibility between these two tools (see the section on ProVerif). This feature is released in CryptoVerif 2.00.

Bruno Blanchet implemented several extensions of CryptoVerif, in particular: 1) reworked the model of Diffie-Hellman key agreements, in particular to account for the absence of public key validation in popular Diffie-Hellman groups like Curve25519, which is used in many modern protocols, 2) support for the proof of indistinguishability between two games given by the user, 3) facilitate the interactive proofs. Program points, used for instance to insert case distinctions, can now be designated as the line that matches a regular expression, instead of using a number. This is much more stable in case the protocol model is slightly modified. Groups of variables can be designated as all variables that match a regular expression. These features are not released yet.

- Participants: Bruno Blanchet and David Cadé
- Contact: Bruno Blanchet
- Publications: Composition Theorems for CryptoVerif and Application to TLS 1.3 - Composition Theorems for CryptoVerif and Application to TLS 1.3 - Proved Implementations of Cryptographic Protocols in the Computational Model - Proved Generation of Implementations from Computationally Secure Protocol Specifications - Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate - Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate - Symbolic and Computational Mechanized Verification of the ARINC823 Avionic Protocols - Automated Verification for Secure Messaging Protocols and Their Implementations: A Symbolic and Computational Approach
- URL: http://cryptoverif.inria.fr/

## 6.3. F*

*FStar*
KEYWORDS: Programming language - Software Verification
FUNCTIONAL DESCRIPTION: F* is a new higher order, effectful programming language (like ML) designed with program verification in mind. Its type system is based on a core that resembles System Fw (hence the name), but is extended with dependent types, refined monadic effects, refinement types, and higher kinds. Together, these features allow expressing precise and compact specifications for programs, including functional correctness properties. The F* type-checker aims to prove that programs meet their specifications using an automated theorem prover (usually Z3) behind the scenes to discharge proof obligations. Programs written in F* can be translated to OCaml, F#, or JavaScript for execution.

- Participants: Antoine Delignat-Lavaud, Catalin Hritcu, Cédric Fournet, Chantal Keller, Karthikeyan Bhargavan and Pierre-Yves Strub
- Contact: Catalin Hritcu
- URL: https://www.fstar-lang.org/

## 6.4. miTLS

KEYWORDS: Cryptographic protocol - Software Verification
FUNCTIONAL DESCRIPTION: miTLS is a verified reference implementation of the TLS protocol. Our code fully supports its wire formats, ciphersuites, sessions and connections, re-handshakes and resumptions, alerts and errors, and data fragmentation, as prescribed in the RFCs, it interoperates with mainstream web browsers and servers. At the same time, our code is carefully structured to enable its modular, automated verification, from its main API down to computational assumptions on its cryptographic algorithms.

- Participants: Alfredo Pironti, Antoine Delignat-Lavaud, Cédric Fournet, Jean-Karim Zinzindohoué, Karthikeyan Bhargavan, Pierre-Yves Strub and Santiago Zanella-Béguelin
- Contact: Karthikeyan Bhargavan
- URL: https://github.com/mitls/mitls-fstar

## 6.5. ProVerif

KEYWORDS: Security - Verification - Cryptographic protocol
FUNCTIONAL DESCRIPTION: ProVerif is an automatic security protocol verifier in the symbolic model (so called Dolev-Yao model). In this model, cryptographic primitives are considered as black boxes. This protocol verifier is based on an abstract representation of the protocol by Horn clauses. Its main features are:

It can verify various security properties (secrecy, authentication, process equivalences).

It can handle many different cryptographic primitives, specified as rewrite rules or as equations.

It can handle an unbounded number of sessions of the protocol (even in parallel) and an unbounded message space.
NEWS OF THE YEAR: Marc Sylvestre extended his interactive simulator of protocols modeled in ProVerif to simulate the semantics of biprocesses, used to prove observational equivalence between two processes. He also made minor improvements to this simulator and to the graphical display of attacks.

Bruno Blanchet modified ProVerif and CryptoVerif to improve the compatibility between these two tools. It is now possible for simple examples to use the same input file with both tools, for instance to try to find attacks in the symbolic model using ProVerif, and if no attack is found, then prove the protocol in the computational model using CryptoVerif. For more complex examples, the differences between the files to provide for each tool are considerably reduced. The cryptographic primitives are specified in distinct libraries, one for ProVerif and one for CryptoVerif, because the assumptions on primitives are very different in the symbolic and computational models. These features are released in ProVerif 2.00.

Vincent Cheval and Bruno Blanchet implemented several extensions of ProVerif: 1) support for integer counters, with incrementation and inequality tests, 2) lemmas and axioms to give intermediate results to ProVerif, which it exploits to help proving subsequent queries, by deriving additional information in the Horn clauses that it uses to perform the proofs, 3) proofs by induction on the length of the trace, by giving as lemma the property to prove, but obviously for strictly shorter traces. These features are not released yet.

- Participants: Bruno Blanchet, Marc Sylvestre and Vincent Cheval
- Contact: Bruno Blanchet
- Publications: Automated reasoning for equivalences in the applied pi calculus with barriers - Automated Reasoning for Equivalences in the Applied Pi Calculus with Barriers - Automated reasoning for equivalences in the applied pi calculus with barriers - Modeling and Verifying Security Protocols with the Applied Pi Calculus and ProVerif - Automatic Verification of Security Protocols in the Symbolic Model: The Verifier ProVerif - Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate - Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate - Automated Verification for Secure Messaging Protocols and Their Implementations: A Symbolic and Computational Approach - Symbolic and Computational Mechanized Verification of the ARINC823 Avionic Protocols - Symbolic and Computational Mechanized Verification of the ARINC823 Avionic Protocols
- URL: http://proverif.inria.fr/

## 6.6. HACL*

*High Assurance Cryptography Library*
KEYWORDS: Cryptography - Software Verification
FUNCTIONAL DESCRIPTION: HACL* is a formally verified cryptographic library in F*, developed by the Prosecco team at Inria Paris in collaboration with Microsoft Research, as part of Project Everest.

HACL stands for High-Assurance Cryptographic Library and its design is inspired by discussions at the HACS series of workshops. The goal of this library is to develop verified C reference implementations for popular cryptographic primitives and to verify them for memory safety, functional correctness, and secret independence.

- Contact: Karthikeyan Bhargavan
- URL: https://github.com/mitls/hacl-star

# 7. New Results

## 7.1. Composition Theorems for CryptoVerif and Application to TLS 1.3

**Participant:** Bruno Blanchet.

We presented composition theorems for security protocols, to compose a key exchange protocol and a symmetric-key protocol that uses the exchanged key. Our results rely on the computational model of cryptography and are stated in the framework of the tool CryptoVerif. They support key exchange protocols that guarantee injective or non-injective authentication. They also allow random oracles shared between the composed protocols. To our knowledge, they are the first composition theorems for key exchange stated for a computational protocol verification tool, and also the first to allow such flexibility.

As a case study, we applied our composition theorems to a proof of TLS 1.3 Draft-18. This work fills a gap in our previous analysis of TLS 1.3 in CryptoVerif [52]. It appears in [31], [39].

## 7.2. Mechanised Cryptographic Proof of the WireGuard VPN Protocol

**Participants:** Benjamin Lipp, Bruno Blanchet, Karthikeyan Bhargavan.

WireGuard is a free and open source Virtual Private Network (VPN) that aims to replace IPsec and OpenVPN. It is based on a new cryptographic protocol derived from the Noise Protocol Framework. We provide the first mechanised cryptographic proof of the protocol underlying WireGuard, using the CryptoVerif proof assistant.

We analyse the entire WireGuard protocol as it is, including transport data messages, in an ACCE-style model. We contribute proofs for correctness, message secrecy, forward secrecy, mutual authentication, session uniqueness, and resistance against key compromise impersonation, identity mis-binding, and replay attacks. We also discusse the strength of the identity hiding provided by WireGuard.

Our work also provides novel theoretical contributions that are reusable beyond WireGuard. First, we extend CryptoVerif to account for the absence of public key validation in popular Diffie-Hellman groups like Curve25519, which is used in many modern protocols including WireGuard. To our knowledge, this is the first mechanised cryptographic proof for any protocol employing such a precise model. Second, we prove several indifferentiability lemmas that are useful to simplify the proofs for sequences of key derivations. This work is under submission.

## 7.3. Meta-F*: Proof automation with SMT, Tactics, and Metaprograms

**Participants:** Guido Martinez, Danel Ahman, Victor Dumitrescu, Nick Giannarakis [Princeton University], Chris Hawblitzel [Microsoft Research], Catalin Hritcu, Monal Narasimhamurthy [University of Colorado Boulder], Zoe Paraskevopoulou [Princeton University], Clément Pit-Claudel [MIT], Jonathan Protzenko [Microsoft Research], Tahina Ramananandro [Microsoft Research], Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research].

We introduced Meta-F* [69], a tactics and metaprogramming framework for the F* program verifier. The main novelty of Meta-F* is allowing to use tactics and metaprogramming to discharge assertions not solvable by SMT, or to just simplify them into well-behaved SMT fragments. Plus, Meta-F* can be used to generate verified code automatically.

Meta-F* is implemented as an F* effect, which, given the powerful effect system of F*, heavily increases code reuse and even enables the lightweight verification of metaprograms. Metaprograms can be either interpreted, or compiled to efficient native code that can be dynamically loaded into the F* type-checker and can interoperate with interpreted code. Evaluation on realistic case studies shows that Meta-F* provides substantial gains in proof development, efficiency, and robustness.

## 7.4. When Good Components Go Bad: Formally Secure Compilation Despite Dynamic Compromise

**Participants:** Carmine Abate, Arthur Azevedo de Amorim [CMU], Roberto Blanco, Ana Nora Evans [University of Virginia], Guglielmo Fachini [Nozomi Networks], Catalin Hritcu, Théo Laurent, Benjamin C. Pierce [University of Pennsylvania], Marco Stronati [Nomadic Labs], Andrew Tolmach [Portland State University].

We proposed a new formal criterion [47] for evaluating secure compilation schemes for unsafe languages, expressing end-to-end security guarantees for software components that may become compromised after encountering undefined behavior—for example, by accessing an array out of bounds.

Our criterion is the first to model dynamic compromise in a system of mutually distrustful components with clearly specified privileges. It articulates how each component should be protected from all the others—in particular, from components that have encountered undefined behavior and become compromised. Each component receives secure compilation guarantees—in particular, its internal invariants are protected from compromised components—up to the point when this component itself becomes compromised, after which we assume an attacker can take complete control and use this component's privileges to attack other components. More precisely, a secure compilation chain must ensure that a dynamically compromised component cannot break the safety properties of the system at the target level any more than an arbitrary attacker-controlled component (with the same interface and privileges, but without undefined behaviors) already could at the source level.

To illustrate the model, we construct a secure compilation chain for a small unsafe language with buffers, procedures, and components, targeting a simple abstract machine with built-in compartmentalization. We give a careful proof (mostly machine-checked in Coq) that this compiler satisfies our secure compilation criterion. Finally, we show that the protection guarantees offered by the compartmentalized abstract machine can be achieved at the machine-code level using either software fault isolation or a tag-based reference monitor.

## 7.5. The Meaning of Memory Safety

**Participants:** Arthur Azevedo de Amorim [CMU], Catalin Hritcu, Benjamin C. Pierce [University of Pennsylvania].

We give a rigorous characterization of what it means for a programming language to be memory safe [51], capturing the intuition that memory safety supports local reasoning about state. We formalize this principle in two ways. First, we show how a small memory-safe language validates a noninterference property: a program can neither affect nor be affected by unreachable parts of the state. Second, we extend separation logic, a proof system for heap-manipulating programs, with a memory-safe variant of its frame rule. The new rule is stronger because it applies even when parts of the program are buggy or malicious, but also weaker because it demands a stricter form of separation between parts of the program state. We also consider a number of pragmatically motivated variations on memory safety and the reasoning principles they support. As an application of our characterization, we evaluate the security of a previously proposed dynamic monitor for memory safety of heap-allocated data.

## 7.6. Recalling a Witness: Foundations and Applications of Monotonic State

**Participants:** Danel Ahman, Cédric Fournet [Microsoft Research], Catalin Hritcu, Kenji Maillard, Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research].

We provide a way to ease the verification of programs whose state evolves monotonically [48]. The main idea is that a property witnessed in a prior state can be soundly recalled in the current state, provided (1) state evolves according to a given preorder, and (2) the property is preserved by this preorder. In many scenarios, such monotonic reasoning yields concise modular proofs, saving the need for explicit program invariants. We distill our approach into the monotonic-state monad, a general yet compact interface for Hoare-style reasoning about monotonic state in a dependently typed language. We prove the soundness of the monotonic-state monad and use it as a unified foundation for reasoning about monotonic state in the F* verification system. Based on this foundation, we build libraries for various mutable data structures like monotonic references and apply these libraries at scale to the verification of several distributed applications.

## 7.7. A Monadic Framework for Relational Verification: Applied to Information Security, Program Equivalence, and Optimizations

**Participants:** Niklas Grimm [Vienna University of Technology], Kenji Maillard, Cédric Fournet [Microsoft Research], Catalin Hritcu, Matteo Maffei [Vienna University of Technology], Jonathan Protzenko [Microsoft Research], Tahina Ramananandro [Microsoft Research], Aseem Rastogi [Microsoft Research], Nikhil Swamy [Microsoft Research], Santiago Zanella-Béguelin [Microsoft Research].

Relational properties describe multiple runs of one or more programs. They characterize many useful notions of security, program refinement, and equivalence for programs with diverse computational effects, and they have received much attention in the recent literature. Rather than developing separate tools for special classes of effects and relational properties, we advocate using a general purpose proof assistant as a unifying framework for the relational verification of effectful programs. The essence of our approach is to model effectful computations using monads and to prove relational properties on their monadic representations, making the most of existing support for reasoning about pure programs [67].

We apply this method in F* and evaluate it by encoding a variety of relational program analyses, including information flow control, program equivalence and refinement at higher order, correctness of program optimizations and game-based cryptographic security. By relying on SMT-based automation, unary weakest preconditions, user-defined effects, and monadic reification, we show that, compared to unary properties, verifying relational properties requires little additional effort from the F* programmer.

## 7.8. A Formal Treatment of Accountable Proxying over TLS

**Participants:** Karthikeyan Bhargavan, Ioana Boureanu [University of Surrey], Antoine Delignat-Lavaud [Microsoft Research], Pierre-Alain Fouque [University of Rennes], Cristina Onete [University of Limoges].

Much of Internet traffic nowadays passes through active proxies, whose role is to inspect, filter, cache, or transform data exchanged between two endpoints. To perform their tasks, such proxies modify channel-securing protocols, like TLS, resulting in serious vulnerabilities. Such problems are exacerbated by the fact that middleboxes are often invisible to one or both endpoints, leading to a lack of accountability. A recent protocol, called mcTLS, pioneered accountability for proxies, which are authorized by the endpoints and given limited read/write permissions to application traffic.

Unfortunately, we show that mcTLS is insecure: the protocol modifies the TLS protocol, exposing it to a new class of middlebox-confusion attacks. Such attacks went unnoticed mainly because mcTLS lacked a formal analysis and security proofs. Hence, our second contribution is to formalize the goal of accountable proxying over secure channels. Third, we propose a provably-secure alternative to soon-to-be-standardized mcTLS: a generic and modular protocol-design that carefully composes generic secure channel-establishment protocols, which we prove secure. Finally, we present a proof-of-concept implementation of our design, instantiated with unmodified TLS 1.3 draft 23, and evaluate its overheads [29].

## 7.9. hacspec: towards verifiable crypto standards

**Participants:** Karthikeyan Bhargavan, Franziskus Kiefer [Mozilla], Pierre-Yves Strub [Ecole Polytechnique].

We designed and published hacspec, a formal specification language for cryptographic primitives. Specifications (specs) written in hacspec are succinct, easy to read and implement, and lend themselves to formal verification using a variety of existing tools. The syntax of hacspec is similar to the pseudocode used in cryptographic standards but is equipped with a static type system and syntax checking tools that can find errors. Specs written in hacspec are executable and can hence be tested against test vectors taken from standards and specified in a common format. Finally, hacspec is designed to be compilable to other formal specification languages like F*, EasyCrypt, Coq, and cryptol, so that it can be used as the basis for formal proofs of functional correctness and cryptographic security using various verification frameworks.

We published a paper presenting the syntax, design, and tool architecture of hacspec. We demonstrated the use of the language to specify popular cryptographic algorithms, and developed preliminary compilers from hacspec to F* and to EasyCrypt. Our eventual goal is to invite authors of cryptographic standards to write their pseudocode in hacspec and to help the formal verification community develop the language and tools that are needed to promote high-assurance cryptographic sofware backed by mathematical proofs. All our code is released publicly on GitHub.

## 7.10. Largest-scale user study of secure messaging and API usage

**Participants:** Francesca Musiani [CNRS], Ksenia Ermoshina [CNRS], Harry Halpin, Iness Ben Guirat [INSAT].

As part of the NEXTLEAP EC project, we engaged in the largest ever user study of secure messaging applications, focusing on typical users as well as "high-risk" users in the Middle East and Ukraine, as well as developers.[41]. This work has been shared with standardization efforts such as the IETF Message Layer Security (MLS) effort in which Inria is participating, as well as W3C standardization of the W3C Web Authentication API. This work helped influence the formal verification of the privacy properties of hardware-based cryptographic authentication, which is a feature needed by many at risk users whose accounts are often the focus of hacks. This work has also led a fundamental inquiry into the social governance of standards and the role of formal verification in the future of standards.[42] As this work is highly interdisciplinary, it has featured collaboration with sociologists at CNRS and interns from INSAT in Tunisia, as well as a lecture series hosted at Centre Pompidou under the direction of Bernard Stiegler and Harry Halpin.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

### *8.1.1. ANR*

*8.1.1.1. AnaStaSec*

Title: Static Analysis for Security Properties (ANR générique 2014.)

Other partners: Inria Paris/EPI Antique, Inria Rennes/EPI Celtique, Airbus Operations SAS, AMOSSYS, CEA-LIST, TrustInSoft

Duration: January 2015 - September 2019.

Coordinator: Jérôme Féret, EPI Antique, Inria Paris (France)

Participant: Bruno Blanchet

Abstract: The project aims at using automated static analysis techniques for verifying security and confidentiality properties of critical avionics software.

*8.1.1.2. AJACS*

Title: AJACS: Analyses of JavaScript Applications: Certification and Security

Other partners: Inria-Rennes/Celtique, Inria-Saclay/Toccata, Inria-Sophia Antipolis/INDES, Imperial College London

Duration: October 2014 - March 2019.

Coordinator: Alan Schmitt, Inria (France)

Participants: Karthikeyan Bhargavan, Bruno Blanchet, Nadim Kobeissi

Abstract: The goal of the AJACS project is to provide strong security and privacy guarantees for web application scripts. To this end, we propose to define a mechanized semantics of the full JavaScript language, the most widely used language for the Web, to develop and prove correct analyses for JavaScript programs, and to design and certify security and privacy enforcement mechanisms.

*8.1.1.3. SafeTLS*

Title: SafeTLS: La sécurisation de l'Internet du futur avec TLS 1.

Other partners: Université Rennes 1, IRMAR, Inria Sophia Antipolis, SGDSN/ANSSI

Duration: October 2016 - September 2020

Coordinator: Pierre-Alain Fouque, Université de Rennes 1 (France)

Participants: Karthikeyan Bhargavan

Abstract: Our project, SafeTLS, addresses the security of both TLS 1.3 and of TLS 1.2 as they are (expected to be) used, in three important ways: (1) A better understanding: We will provide a better understanding of how TLS 1.2 and 1.3 are used in real-world applications; (2) Empowering clients: By developing a tool that will show clients the quality of their TLS connection and inform them of potential security and privacy risks; (3) Analyzing implementations: We will analyze the soundness of current TLS 1.2 implementations and use automated verification to provide a backbone of a secure TLS 1.3 implementation.

*8.1.1.4. TECAP*

Title: TECAP: Protocol Analysis - Combining Existing Tools (ANR générique 2017.)

Other partners: Inria Nancy/EPI PESTO, Inria Sophia Antipolis/EPI MARELLE, IRISA, LIX, LSV - ENS Cachan.

Duration: January 2018 - December 20

Coordinator: Vincent Cheval, EPI PESTO, Inria Nancy (France)

Participants: Bruno Blanchet, Benjamin Lipp

Abstract: A large variety of automated verification tools have been developed to prove or find attacks on security protocols. These tools differ in their scope, degree of automation, and attacker models. The aim of this project is to get the best of all these tools, meaning, on the one hand, to improve the theory and implementations of each individual tool towards the strengths of the others and, on the other hand, build bridges that allow the cooperations of the methods/tools. We will focus in this project on the tools CryptoVerif, EasyCrypt, Scary, ProVerif, Tamarin, AKiSs and APTE.

## 8.2. European Initiatives

### *8.2.1. FP7 & H2020 Projects*

*8.2.1.1. ERC Consolidator Grant: CIRCUS*

Title: CIRCUS: An end-to-end verification architecture for building Certified Implementations of Robust, Cryptographically Secure web applications

Duration: April 2016 - March 2021

Coordinator: Karthikeyan Bhargavan, Inria

The security of modern web applications depends on a variety of critical components including cryptographic libraries, Transport Layer Security (TLS), browser security mechanisms, and single sign-on protocols. Although these components are widely used, their security guarantees remain poorly understood, leading to subtle bugs and frequent attacks. Rather than fixing one attack at a time, we advocate the use of formal security verification to identify and eliminate entire classes of vulnerabilities in one go.

CIRCUS proposes to take on this challenge, by verifying the end-to-end security of web applications running in mainstream software. The key idea is to identify the core security components of web browsers and servers and replace them by rigorously verified components that offer the same functionality but with robust security guarantees.

*8.2.1.2. ERC Starting Grant: SECOMP*

Title: SECOMP: Efficient Formally Secure Compilers to a Tagged Architecture

Duration: Jan 2017 - December 2021

Coordinator: Catalin Hritcu, Inria

Abstract: The SECOMP project is aimed at leveraging emerging hardware capabilities for fine-grained protection to build the first, efficient secure compilation chains for realistic low-level programming languages (the C language, and Low* a safe subset of C embedded in F* for verification). These compilation chains will provide a secure semantics for all programs and will ensure that high-level abstractions cannot be violated even when interacting with untrusted low-level code. To achieve this level of security without sacrificing efficiency, our secure compilation chains target a tagged architecture, which associates a metadata tag to each word and efficiently propagates and checks tags according to software-defined rules. We will use property-based testing and formal verification to provide high confidence that our compilers are indeed secure.

*8.2.1.3. NEXTLEAP*

Title: NEXTLEAP: NEXT generation Legal Encryption And Privacy

Programm: H2020

Duration: January 2016 - December 2018

Coordinator: Harry Halpin, Inria

Other partners: IMDEA, University College London, CNRS, IRI, and Merlinux

Abstract: NEXTLEAP aims to create, validate, and deploy protocols that can serve as pillars for a secure, trust-worthy, and privacy-respecting Internet. For this purpose NEXTLEAP will develop an interdisciplinary study of decentralisation that provides the basis on which these protocols cann be designed, working with sociologists to understand user needs. The modular specification of decentralized protocols, implemented as verified open-source software modules, will be done for both privacy-preserving secure federated identity as well as decentralized secure messaging services that hide metadata (e.g., who, when, how often, etc.).

# 8.3. International Initiatives

## *8.3.1. Inria International Partners*

*8.3.1.1. Informal International Partners*

We have a range of long- and short-term collaborations with various universities and research labs. We summarize them by project:

- TLS analysis: Microsoft Research (Cambridge), Mozilla, University of Rennes
- F*: Microsoft Research (Redmond, Cambridge, Bangalore), MSR-Inria, CMU, MIT, University of Ljubljana, Nomadic Labs, Zen Protocol, Princeton University
- SECOMP: MPI-SWS, CISPA, Stanford University, CMU, University of Pennsylvania, Portland State University, University of Virginia, University of Iai
- Micro-Policies: University of Pennsylvania, Portland State University, MIT, Draper Labs, Dover Microsystems

### *8.3.2. Participation in Other International Programs*

*8.3.2.1. SSITH/HOPE*

Title: Advanced New Hardware Optimized for Policy Enforcement, A New HOPE

Program: DARPA SSITH

Duration: December 2017 - February 2021

Coordinator: Charles Stark Draper Laboratory

Other Participants: Inria Paris, University of Pennsylvania, MIT, Portland State University, Dover Microsystems, DornerWorks

Participants from Inria Prosecco: Catalin Hritcu, Roberto Blanco, Jérémy Thibault

Abstract: A New HOPE builds on results from the Inherently Secure Processor (ISP) project that has been internally funded at Draper. Recent architectural improvements decouple the tagged architecture from the processor pipeline to improve performance and flexibility for new processors. HOPE securely maintains metadata for each word in application memory and checks every instruction against a set of installed security policies. The HOPE security architecture exposes tunable parameters that support Performance, Power, Area, Software compatibility and Security (PPASS) search space exploration. Flexible software-defined security policies cover all 7 SSITH CWE vulnerability classes, and policies can be tuned to meet PPASS requirements; for example, one can trade granularity of security checks against performance using different policy configurations. HOPE will design and formalize a new high-level domain-specific language (DSL) for defining security policies, based on previous research and on extensive experience with previous policy languages. HOPE will formally verify that installed security policies satisfy system-wide security requirements. A secure boot process enables policies to be securely updated on deployed HOPE systems. Security policies can adapt based on previously detected attacks. Over the multi-year, multi-million dollar Draper ISP project, the tagged security architecture approach has evolved from early prototypes based on results from the DARPA CRASH program towards easier integration with external designs, and is better able to scale from micro to server class implementations. A New HOPE team is led by Draper and includes faculty from University of Pennsylvania (Penn), Portland State University (PSU), Inria, and MIT, as well as industry collaborators from DornerWorks and Dover Microsystems. In addition to Draper's in-house expertise in hardware design, cyber-security (defensive and offensive, hardware and software) and formal methods, the HOPE team includes experts from all domains relevant to SSITH, including (a) computer architecture: DeHon (Penn), Shrobe (MIT); (b) formal methods including programming languages and security: Pierce (Penn), Tolmach (PSU), Hritcu (Inria); and (c) operating system integration (DornerWorks). Dover Microsystems is a spin-out from Draper that will commercialize concepts from the Draper ISP project.

*8.3.2.2. Everest Expedition*

Program: Microsoft Expedition and MSR-Inria Collaborative Research Project

Expedition Participants: Microsoft Research (Cambridge, Redmond, Bangalore), Inria, MSR-Inria, CMU, University of Edinburgh

Duration of current MSR-Inria Project: October 2017 – October 2020

Participants from Inria Prosecco: Karthikeyan Bhargavan, Catalin Hritcu, Danel Ahman, Benjamin Beurdouche, Victor Dumitrescu, Nadim Kobeissi, Théo Laurent, Guido Martínez, Denis Merigoux, Marina Polubelova, Jean-Karim Zinzindohoué

Participants from other Inria teams: David Pichardie (Celtique), Jean-Pierre Talpin (TEA)

Abstract: The HTTPS ecosystem (HTTPS and TLS protocols, X.509 public key infrastructure, crypto algorithms) is the foundation on which Internet security is built. Unfortunately, this ecosystem is brittle, with headline-grabbing attacks such as FREAK and LogJam and emergency patches many times a year.

Project Everest addresses this problem by constructing a high-performance, standards-compliant, formally verified implementation of components in HTTPS ecosystem, including TLS, the main protocol at the heart of HTTPS, as well as the main underlying cryptographic algorithms such as AES, SHA2 or X25519.

At the TLS level, for instance, we are developing new implementations of existing and forthcoming protocol standards and formally proving, by reduction to cryptographic assumptions on their core algorithms, that our implementations provide a secure-channel abstraction between the communicating endpoints. Implementations of the core algorithms themselves are also verified, producing performant portable C code or highly optimized assembly language.

We aim for our verified components to be drop-in replacements suitable for use in mainstream web browsers, servers, and other popular tools and are actively working with the community at large to improve the ecosystem.

https://project-everest.github.io

## 8.4. International Research Visitors

### *8.4.1. Visits of International Scientists*

- Amal Ahmed (Northeastern University, USA) joined Inria as a Visiting Professor from September 2017 to July 2018; she gave a seminar on "Compositional Compiler Verification for a Multi-Language World".
- Aaron Weiss (Northeastern University, USA) joined Inria as a Visiting Scientist from September 2017 to July 2018; he gave a seminar on "Rust Distilled: An Expressive Tower of Languages"
- Justin Hsu (University of Wisconsin–Madison, USA) visited Prosecco on 26 January 2018 and gave a talk entitled "From Couplings to Probabilistic Relational Program Logics"
- Deepak Garg (MPI-SWS, Germany) visited Prosecco on 21 February and 6 December 2018
- Marco Patrignani (CISPA, Germany) visited Prosecco on 21 February 2018
- Arthur Azevedo de Amorim (CMU) visited Prosecco on 10–13 April 2018 and gave a seminar on "The Meaning of Memory Safety"
- Prasad Naldurg (IBM Research, India) joined Prosecco as a Visiting Researcher from May 2018; he gave a Prosecco seminar on "Encrypted Analytics: Computing directly on encrypted databases"
- Vincent Gramoli (NICTA/Data61-CSIRO and University of Sydney, Australia) visited Prosecco on 27 June 2018 and gave a seminar on "The Red Belly Blockchain: Speed, Security, Scalability"
- Éric Tanter (University of Chile) joined Prosecco as Visiting Professor from July 2018 to February 2019; he gave a Prosecco seminar on "Gradual Parametricity, Revisited" and many other talks
- Andrew Tolmach (Portland State University, USA) visited Prosecco on 2–4 July 2018
- Ilya Sergey (University College London, UK) visited Prosecco on 5 September 2018 and gave a seminar on "Deductive Synthesis of Programs that Alter Data Structures"
- Jonathan Aldrich (CMU, USA) visited Prosecco on 22–26 November 2018 and gave a seminar on "Object Capabilities, Effects, and Abstraction"
- tefan Ciobâcă (University of Iai, Romania) visited Prosecco on 3–7 December 2018
- Amin Timany (KU Leuven, Belgium) visited Prosecco on 3–7 December 2018
- Cédric Fournet (Microsoft Research, UK) has visited Prosecco on various occasions
- Jonathan Protzenko (Microsoft Research, USA) has visited Prosecco on various occasions

#### *8.4.1.1. Internships*

- Benjamin Lipp (Karlsruhe Institute of Technology, Germany): from Dec 2017 to May 2018 – advised by Bruno Blanchet and Karthik Bhargavan

- Carmine Abate (University of Trento, Italy): from Dec 2017 to May 2018 – advised by Catalin Hritcu
- Jérémy Thibault (ENS Rennes, France): from Feb to Jul 2018 – advised by Catalin Hritcu
- Florian Groult (University of Orleans, France): from Apr to Oct 2018 – advised by Catalin Hritcu
- Guido Martinez (CIFASIS-CONICET Rosario, Argentina): from Sep to December 2018 – advised by Catalin Hritcu
- Elizabeth Labrada Deniz (University of Chile): from Oct 2018 to January 2019 – advised by Éric Tanter and Catalin Hritcu
- Iness Ben Guirat (INSAT): from August 2018 to January 2019 – advised by Harry Halpin

### *8.4.2. Visits to International Teams*

- Catalin Hritcu, Danel Ahman, and Victor Dumitrescu visited Microsoft Research (Redmond, USA) on 5–25 March 2018
- Catalin Hritcu, Carmine Abate, and Jérémy Thibault visited the MPI-SWS (Saarbrücken, Germany) on 27–28 March 2018
- Catalin Hritcu visited Draper Labs (Cambridge, MA, USA) on 30 May 2018
- Karthikeyan Bhargavan, Catalin Hritcu, Danel Ahman, Benjamin Beurdouche, Victor Dumitrescu, Guido Martínez, Denis Merigoux, and Marina Polubelova visited Microsoft Research (Cambridge, UK) for Everest "All-Hands" meeting
- Harry Halpin visited the NEXTLEAP team meeting (Lausanne, Switzerland) on 15–17th of January.
- Harry Halpin visited the NEXTLEAP team meeting (Freibourg, Germany) on 21–22nd of November.
- Harry Halpin visited the final PANORAMIX team meeting (Athens, Greece) on 24–25th of September.

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Organisation*

*9.1.1.1. General Chair, Scientific Chair*

- Catalin Hritcu and Amal Ahmed co-organized a Dagstuhl Seminar on Secure Compilation (18201)
- Harry Halpin and Bart Preneel co-organized the ECRYPT-CSA workshop on Crypto Policies (22-23 January 2018) in Brussels, Belgium.

*9.1.1.2. Member of the Organizing Committees*

- Catalin Hritcu and Amal Ahmed were organizers for PriSC 2018 and the upcoming PriSC 2019

### *9.1.2. Scientific Events Selection*

*9.1.2.1. Chair of Conference Program Committees*

- Catalin Hritcu was the PC chair for the 2nd Workshop on Principles of Secure Compilation (PriSC) at POPL 2018
- Harry Halpin was General Chair of the 1st Workshop on the Decentralization of Governance at INSCI 2018

*9.1.2.2. Member of the Conference Program Committees*

- Bruno Blanchet was PC member of RESSI 2018 (*Rendez-vous de la Recherche et de l'Enseignement de la Sécurité des Systèmes d'Information*).
- Catalin Hritcu was PC member of EuroS&P 2018, ESOP 2018, and CCS 2018
- Karthikeyan Bhargavan was PC member of IEEE S&P 2018, ACM CCS 2018, and POST 2018
- Harry Halpin was PC Member of SSR 2018, ACM WWW 2018, and ISWC 2018.

*9.1.2.3. Reviewer*

- Catalin Hritcu served as a reviewer for the Journal of Automated Reasoning (JAR)

### *9.1.3. Journal*

*9.1.3.1. Member of the Editorial Boards*

Associate Editor

– of the *International Journal of Applied Cryptography (IJACT)* – Inderscience Publishers: Bruno Blanchet

### *9.1.4. Invited Talks*

- Catalin Hritcu gave an Invited Keynote talk at the Working Formal Methods Symposium (FROM) in June 2018
- Catalin Hritcu gave invited talks at Nomadic Labs (Tezos), IRIF Verification Seminar (Paris 7), and SoSySec seminar (IRISA Rennes)
- Karthikeyan Bhargavan gave invited talks at Security Standardization Research (SSR 2018), Formal Methods and Tools for Security (FMATS 2018), Crypto Welcomes TLS 1.3 (CWTLS), and the annual GDR Securité meeting.
- Harry Halpin gave invited talks at the EPFL Summer Research Institute in July 2018, the Web 3.0 Summit in October 2018, and Binance Labs in December 2018.

### *9.1.5. Leadership within the Scientific Community*

- Catalin Hritcu served as the Artifact Evaluation Co-Chair for POPL 2018 and POPL 2019

### *9.1.6. Scientific Expertise*

- Bruno Blanchet is a member of the specialized temporary scientific committee of ANSM (*Agence nationale de sécurité du médicament et des produits de santé*), on the cybersecurity of software medical devices.
- Bruno Blanchet participated to a review of the code of the Tezos blockchain by the Inria Foundation (March–May 2018).
- Harry Halpin participated as a member of the advisory board to the PANORAMIX EC H2020 project (2018).

### *9.1.7. Research Administration*

- Bruno Blanchet was co-president of the Inria hiring committee for PhD, post-docs, and *délégations* (*Commision des Emplois Scientifiques*, CES).
- Bruno Blanchet was representative of Inria Paris at the DIM RFSI (*Domaine d'Intérêt Majeur, Réseau Francilien en Sciences Informatiques*).

## 9.2. Teaching - Supervision - Juries

### *9.2.1. Teaching*

- Master: Bruno Blanchet, Cryptographic protocols: formal and computational proofs, 18h equivalent TD, master M2 MPRI, université Paris VII, France

- Master: Karthikeyan Bhargavan, Cryptographic protocols: formal and computational proofs, 18h equivalent TD, master M2 MPRI, université Paris VII, France
- Master: Karthikeyan Bhargavan, Network Protocol Safety and Security, 18h equivalent TD, ACN master, Telecom ParisTech
- PhD: Formally Secure Compartmentalizing Compilation course at International School on Foundations of Security Analysis and Design (FOSAD), 27-28 August, 2018, Bertinoro, Italy
- PhD: Program Verification with F* course at EPIT 2018 Software Verification Spring School, 7-11 May, 2018, Aussois, France
- PhD: Attacks and Automated Tools, BIU winter school on cryptography, 11-15 February, 2018, Tel Aviv, Israel
- PhD: Crypto standards for the Internet and Web. ECRYPT-CSA School on Societal Aspects of Cryptology and on Business and Innovation in Crypto. 7-9 January. Zurich, Switzerland.
- PhD: Mix networking, ECRYPT Summer School, ECRYPT-NET School on Integrating Advanced Cryptography with Applications, 16-21 September 2018, Kos, Greece.

### *9.2.2. Supervision*

- PhD: Jean Karim Zinzindohoue, Secure, Fast and Verified Cryptographic Applications: A Scalable Approach [13], ENS Paris, defended on July 3, 2018, supervised by Karthikeyan Bhargavan.
- PhD: Nadim Kobeissi, Formal Verification for Real-World Cryptographic Protocols and Implementations [12], ENS Paris, defended on December 10, 2018, supervised by Karthikeyan Bhargavan and Bruno Blanchet.
- PhD in progress: Benjamin Beurdouche, on verified cryptographic protocol implementations, ENS Paris, since October 2016, supervised by Karthikeyan Bhargavan.
- PhD in progress: Marina Polubelova, on verified post-quantum cryptography, PSL Paris, since October 2017, supervised by Karthikeyan Bhargavan.
- PhD in progress: Natalia Kulatova, on verified secure hardware APIs, PSL Paris, since October 2017, supervised by Karthikeyan Bhargavan.
- PhD in progress: Denis Merigoux, on verified RUST applications, PSL Paris, since November 2017, supervised by Karthikeyan Bhargavan.
- PhD in progress: Benjamin Lipp, On Mechanised Cryptographic Proofs of Protocols and their Link with Verified Implementations, ENS Paris, since October 2018, supervised by Bruno Blanchet and Karthikeyan Bhargavan.
- PhD in progress: Kenji Maillard, on Semantic Foundations for F*, started January 2017, supervised by Catalin Hritcu and Karthikeyan Bhargavan
- PhD in progress: Carmine Abate, The Formal Foundations of Secure Compilation, since June 2018, advised by Catalin Hritcu and Bruno Blanchet
- PhD in progress: Jérémy Thibault, Secure Compartmentalizing Compilation to a Tagged Architecture, from August 2018, advised by Catalin Hritcu and Bruno Blanchet
- PhD in progress: Guido Martínez (CIFASIS-CONICET Rosario), Metatheory for Semi-Automatic Verification of Effectful Programs, from April 2017, advised by Mauro Jaskelioff (CIFASIS-CONICET Rosario) and Catalin Hritcu

### *9.2.3. Juries*

- Karthikeyan Bhargavan participated in the PhD jury of Daniel Fett at University of Stuttgart.
- Harry Halpin participated in the PhD jury of Joseph Raad at University Paris-Saclay.

## 9.3. Popularization

### *9.3.1. Internal or external Inria responsibilities*

- Bruno Blanchet was co-president of the Inria hiring committee for PhD, post-docs, and *délégations* (*Commision des Emplois Scientifiques*, CES).
- Bruno Blanchet was representative of Inria Paris at the DIM RFSI (*Domaine d'Intérêt Majeur, Réseau Francilien en Sciences Informatiques*).

### *9.3.2. Interventions*

- Karthikeyan Bhargavan was a panelist at the Cloudflare Internet Summit in London, June 14, 2018.
- Harry Halpin was a panelist at the World Digital Asset Summit in San Fransisco, USA, December 10, 2018.

# 10. Bibliography

## Major publications by the team in recent years

[1] M. ABADI, B. BLANCHET, C. FOURNET.*The Applied Pi Calculus: Mobile Values, New Names, and Secure Communication*, in "Journal of the ACM (JACM)", October 2017, vol. 65, n° 1, p. 1 - 103 [*DOI :* 10.1145/3127586], https://hal.inria.fr/hal-01636616

[2] C. ABATE, A. AZEVEDO DE AMORIM, R. BLANCO, A. N. EVANS, G. FACHINI, C. HRIŢCU, T. LAURENT, B. C. PIERCE, M. STRONATI, A. TOLMACH.*When Good Components Go Bad: Formally Secure Compilation Despite Dynamic Compromise*, in "25th ACM Conference on Computer and Communications Security (CCS)", Toronto, Canada, ACM, October 2018, p. 1351–1368, https://arxiv.org/abs/1802.00588 [*DOI :* 10.1145/3243734.3243745], https://hal.archives-ouvertes.fr/hal-01949202

[3] A. AZEVEDO DE AMORIM, M. DÉNÈS, N. GIANNARAKIS, C. HRIŢCU, B. C. PIERCE, A. SPECTOR-ZABUSKY, A. TOLMACH.*Micro-Policies: Formally Verified, Tag-Based Security Monitors*, in "36th IEEE Symposium on Security and Privacy (Oakland S&P)", IEEE Computer Society, May 2015, p. 813–830 [*DOI :* 10.1109/SP.2015.55], https://hal.inria.fr/hal-01265666

[4] K. BHARGAVAN, B. BLANCHET, N. KOBEISSI.*Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate*, in "38th IEEE Symposium on Security and Privacy", San Jose, United States, May 2017, p. 483 - 502 [*DOI :* 10.1109/SP.2017.26], https://hal.inria.fr/hal-01575920

[5] K. BHARGAVAN, A. DELIGNAT-LAVAUD, C. FOURNET, A. PIRONTI, P.-Y. STRUB.*Triple Handshakes and Cookie Cutters: Breaking and Fixing Authentication over TLS*, in "IEEE Symposium on Security and Privacy (Oakland)", 2014, p. 98–113, https://hal.inria.fr/hal-01102259

[6] B. BLANCHET.*Modeling and Verifying Security Protocols with the Applied Pi Calculus and ProVerif*, in "Foundations and Trends in Privacy and Security", October 2016, vol. 1, n° 1–2, p. 1–135, https://hal.inria.fr/hal-01423760

[7] M. ISAAKIDIS, H. HALPIN, G. DANEZIS.*UnlimitID: Privacy-Preserving Federated Identity Management Using Algebraic MACs*, in "Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society", New York, NY, USA, WPES '16, ACM, 2016, p. 139–142 [*DOI :* 10.1145/2994620.2994637], https://hal.inria.fr/hal-01426847

[8] N. KOBEISSI, K. BHARGAVAN, B. BLANCHET.*Automated Verification for Secure Messaging Protocols and Their Implementations: A Symbolic and Computational Approach*, in "2nd IEEE European Symposium on Security and Privacy", Paris, France, April 2017, p. 435 - 450 [*DOI :* 10.1109/EUROSP.2017.38], https://hal.inria.fr/hal-01575923

[9] N. SWAMY, C. HRITCU, C. KELLER, A. RASTOGI, A. DELIGNAT-LAVAUD, S. FOREST, K. BHARGAVAN, C. FOURNET, P.-Y. STRUB, M. KOHLWEISS, J.-K. ZINZINDOHOUÉ, S. ZANELLA-BÉGUELIN.*Dependent Types and Multi-Monadic Effects in F**, in "43rd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)", ACM, January 2016, p. 256-270, https://hal.inria.fr/hal-01265793

[10] J.-K. ZINZINDOHOUÉ, K. BHARGAVAN, J. PROTZENKO, B. BEURDOUCHE.*HACL*: A Verified Modern Cryptographic Library*, in "Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017", 2017, p. 1789–1806, https://hal.inria.fr/hal-01588421

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] C. HRIŢCU.*The Quest for Formally Secure Compartmentalizing Compilation*, ENS Paris ; PSL Research University, January 2019, Habilitation à diriger des recherches, https://tel.archives-ouvertes.fr/tel-01995823

[12] N. KOBEISSI.*Formal Verification for Real-World Cryptographic Protocols and Implementations*, Inria Paris ; Ecole Normale Supérieure de Paris - ENS Paris, December 2018, https://hal.inria.fr/tel-01950884

[13] J.-K. ZINZINDOHOUÉ.*Secure, fast and verified cryptographic applications: a scalable approach*, Université de recherche Paris Sciences Lettres – PSL Research University, July 2018, https://hal.inria.fr/tel-01981380

### Articles in International Peer-Reviewed Journal

[14] D. ADRIAN, K. BHARGAVAN, Z. DURUMERIC, P. GAUDRY, M. GREEN, J. A. HALDERMAN, N. HENINGER, D. SPRINGALL, E. THOMÉ, L. VALENTA, B. VANDERSLOOT, E. WUSTROW, S. ZANELLA-BÉGUELIN, P. ZIMMERMANN.*Imperfect forward secrecy: How Diffie-Hellman fails in practice*, in "Communications of the ACM", December 2018, vol. 62, n$^{o}$ 1, p. 106-114 [*DOI :* 10.1145/3292035], https://hal.inria.fr/hal-01982426

[15] D. AHMAN.*Handling Fibred Algebraic Effects*, in "Proceedings of the ACM on Programming Languages", January 2018, vol. 2, n$^{o}$ POPL [*DOI :* 10.1145/3158095], https://hal.archives-ouvertes.fr/hal-01672734

[16] D. AHMAN, C. FOURNET, C. HRIŢCU, K. MAILLARD, A. RASTOGI, N. SWAMY.*Recalling a Witness: Foundations and Applications of Monotonic State*, in "Proceedings of the ACM on Programming Languages", January 2018, vol. 2, n$^{o}$ POPL, https://arxiv.org/abs/1707.02466 [*DOI :* 10.1145/3158153], https://hal.archives-ouvertes.fr/hal-01672733

[17] B. BLANCHET, B. SMYTH.*Automated reasoning for equivalences in the applied pi calculus with barriers*, in "Journal of Computer Security", 2018, vol. 26, n$^{o}$ 3, p. 367 - 422 [*DOI :* 10.3233/JCS-171013], https://hal.inria.fr/hal-01947972

[18] W. J. BOWMAN, Y. CONG, N. RIOUX, A. AHMED.*Type-Preserving CPS Translation of* $\Sigma$ *and* $\Pi$ *Types is Not Not Possible*, in "Proceedings of the ACM on Programming Languages", January 2018, vol. 2, n$^{o}$ POPL [*DOI :* 10.1145/3158110], https://hal.archives-ouvertes.fr/hal-01672735

[19] O. FLÜCKIGER, G. SCHERER, M.-H. YEE, A. GOEL, A. AHMED, J. VITEK.*Correctness of Speculative Optimizations with Dynamic Deoptimization*, in "Proceedings of the ACM on Programming Languages",

2018, vol. 2, n$^{o}$ POPL, https://arxiv.org/abs/1711.03050 [*DOI :* 10.1145/3158137], https://hal.inria.fr/hal-01646765

[20] M. NEW, A. AHMED.*Graduality from embedding-projection pairs*, in "Proceedings of the ACM on Programming Languages", July 2018, vol. 2, n$^{o}$ ICFP, p. 1-30, https://arxiv.org/abs/1807.02786 [*DOI :* 10.1145/3236768], https://hal.archives-ouvertes.fr/hal-01949209

[21] N. TABAREAU, É. TANTER, M. SOZEAU.*Equivalences for Free: Univalent Parametricity for Effective Transport*, in "Proceedings of the ACM on Programming Languages", September 2018, p. 1-29 [*DOI :* 10.1145/3234615], https://hal.inria.fr/hal-01559073

[22] M. TORO, R. GARCIA, É. TANTER.*Type-Driven Gradual Security with References*, in "ACM Transactions on Programming Languages and Systems (TOPLAS)", December 2018, vol. 40, n$^{o}$ 4, p. 1-55 [*DOI :* 10.1145/3229061], https://hal.archives-ouvertes.fr/hal-01957581

[23] M. TORO, E. LABRADA, É. TANTER.*Gradual Parametricity, Revisited*, in "Proceedings of the ACM on Programming Languages", 2018, vol. 3, n$^{o}$ POPL, https://arxiv.org/abs/1807.04596 [*DOI :* 10.1145/3290330], https://hal.archives-ouvertes.fr/hal-01960553

[24] N. VAZOU, É. TANTER, D. VAN HORN.*Gradual liquid type inference*, in "Proceedings of the ACM on Programming Languages", October 2018, vol. 2, n$^{o}$ OOPSLA, p. 1-25, https://arxiv.org/abs/1807.02132 [*DOI :* 10.1145/3276502], https://hal.archives-ouvertes.fr/hal-01949207

### International Conferences with Proceedings

[25] C. ABATE, A. AZEVEDO DE AMORIM, R. BLANCO, A. N. EVANS, G. FACHINI, C. HRIŢCU, T. LAURENT, B. C. PIERCE, M. STRONATI, A. TOLMACH.*When Good Components Go Bad: Formally Secure Compilation Despite Dynamic Compromise*, in "25th ACM Conference on Computer and Communications Security (CCS)", Toronto, Canada, ACM, October 2018, p. 1351–1368, https://arxiv.org/abs/1802.00588 [*DOI :* 10.1145/3243734.3243745], https://hal.archives-ouvertes.fr/hal-01949202

[26] A. AZEVEDO DE AMORIM, C. HRIŢCU, B. C. PIERCE.*The Meaning of Memory Safety*, in "7th International Conference on Principles of Security and Trust (POST)", Thessaloniki, Greece, April 2018, p. 79–105, https://arxiv.org/abs/1705.07354 [*DOI :* 10.1007/978-3-319-89722-6_4], https://hal.archives-ouvertes.fr/hal-01949201

[27] D. BAELDE, A. LICK, S. SCHMITZ.*A Hypersequent Calculus with Clusters for Linear Frames*, in "Twefth Conference on Advances in Modal Logic", Bern, Switzerland, G. BEZHANISHVILI, G. D'AGOSTINO, G. METCALFE, T. STUDE (editors), Advances in Modal Logic, College Publications, July 2018, vol. 12, p. 36–55, https://hal.inria.fr/hal-01756126

[28] D. BAELDE, A. LICK, S. SCHMITZ.*A Hypersequent Calculus with Clusters for Tense Logic over Ordinals*, in "38th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science", Ahmedabad, India, S. GANGULY, P. PANDYA (editors), Leibniz International Proceedings in Informatics, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2018, vol. 122, p. 15:1–15:19 [*DOI :* 10.4230/LIPICS.FSTTCS.2018.15], https://hal.inria.fr/hal-01852077

[29] K. BHARGAVAN, I. BOUREANU, A. DELIGNAT-LAVAUD, P.-A. FOUQUE, C. ONETE.*A Formal Treatment of Accountable Proxying over TLS*, in "SP 2018 - IEEE Symposium on Security and Privacy", San Francisco, United States, May 2018, https://hal.inria.fr/hal-01948722

[30] K. BHARGAVAN, F. KIEFER, P.-Y. STRUB.*hacspec: Towards Verifiable Crypto Standards*, in "Security Standardisation Research. SSR 2018", Darmstadt, Germany, November 2018, p. 1-20 [*DOI :* 10.1007/978-3-030-04762-7_1], https://hal.inria.fr/hal-01967342

[31] B. BLANCHET.*Composition Theorems for CryptoVerif and Application to TLS 1.3*, in "31st IEEE Computer Security Foundations Symposium (CSF'18)", Oxford, United Kingdom, July 2018 [*DOI :* 10.1109/CSF.2018.00009], https://hal.inria.fr/hal-01947959

[32] W. BOWMAN, A. AHMED.*Typed closure conversion for the calculus of constructions*, in "PLDI'18 - 39th ACM SIGPLAN Conference on Programming Language Design and Implementation", Philadelphia, PA, United States, June 2018, https://arxiv.org/abs/1808.04006 [*DOI :* 10.1145/3296979.3192372], https://hal.archives-ouvertes.fr/hal-01949211

[33] N. GRIMM, K. MAILLARD, C. FOURNET, C. HRIŢCU, M. MAFFEI, J. PROTZENKO, T. RAMANANANDRO, A. RASTOGI, N. SWAMY, S. ZANELLA-BÉGUELIN.*A Monadic Framework for Relational Verification: Applied to Information Security, Program Equivalence, and Optimizations*, in "7th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP)", Los Angeles, United States, ACM, January 2018, p. 130–145, https://arxiv.org/abs/1703.00055 [*DOI :* 10.1145/3167090], https://hal.archives-ouvertes.fr/hal-01672703

[34] N. KOBEISSI, N. KULATOVA.*Ledger Design Language: Designing and Deploying Formally Verified Public Ledgers*, in "Workshop on Security Protocol Implementations: Development and Analysis", London, United Kingdom, April 2018, https://hal.inria.fr/hal-01948971

[35] G. SCHERER, M. NEW, N. RIOUX, A. AHMED.*FabULous Interoperability for ML and a Linear Language*, in "International Conference on Foundations of Software Science and Computation Structures (FoSSaCS)", Thessaloniki, Greece, C. BAIE, U. D. LAGO (editors), FabOpen image in new windowous Interoperability for ML and a Linear Language, Springer, April 2018, vol. LNCS - Lecture Notes in Computer Science, n$^{o}$ 10803, https://arxiv.org/abs/1707.04984 [*DOI :* 10.1007/978-3-319-89366-2_8], https://hal.inria.fr/hal-01929158

### Conferences without Proceedings

[36] N. KOBEISSI, K. BHARGAVAN.*Noise Explorer: Fully Automated Modeling and Verification for Arbitrary Noise Protocols*, in "RWC 2019 - Real World Cryptography Symposium", San Jose, United States, January 2019, https://hal.inria.fr/hal-01948964

[37] K. MAILLARD, É. MIQUEY, X. MONTILLET, G. MUNCH-MACCAGNONI, G. SCHERER.*A preview of a tutorial on L (polarized $\mu\mu$-tilde)*, in "HOPE 2018 - 7th ACM SIGPLAN Workshop on Higher-Order Programming with Effects", St. Louis, United States, September 2018, https://hal.inria.fr/hal-01992294

[38] G. MARTÍNEZ, D. AHMAN, V. DUMITRESCU, N. GIANNARAKIS, C. HAWBLITZEL, C. HRIŢCU, M. NARASIMHAMURTHY, Z. PARASKEVOPOULOU, C. PIT-CLAUDEL, J. PROTZENKO, T. RAMANANANDRO, A. RASTOGI, N. SWAMY.*Meta-F*: Proof automation with SMT, Tactics, and Metaprograms*, in "ESOP'19 - European Symposium on Programming", Prague, Czech Republic, April 2019, https://arxiv.org/abs/1803.06547 , https://hal.archives-ouvertes.fr/hal-01995376

### Research Reports

[39] B. BLANCHET.*Composition Theorems for CryptoVerif and Application to TLS 1.3*, Inria Paris, April 2018, n^o RR-9171, 67, https://hal.inria.fr/hal-01764527

### Other Publications

[40] D. BAELDE, A. LICK, S. SCHMITZ.*Decidable XPath Fragments in the Real World*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01852475

[41] H. HALPIN, K. ERMOSHINA, F. MUSIANI.*Co-ordinating Developers and High-Risk Users of Privacy-Enhanced Secure Messaging Protocols*, November 2018, SSR 2018 - Security Standardisation Research Conference, https://hal.inria.fr/hal-01966560

[42] H. HALPIN.*Decentralizing the Social Web*, October 2018, INSCI'2018- 5th International conference 'Internet Science', https://hal.inria.fr/hal-01966561

[43] N. KOBEISSI.*Capsule: A Protocol for Secure Collaborative Document Editing*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01948967

## References in notes

[44] M. ABADI, B. BLANCHET.*Analyzing Security Protocols with Secrecy Types and Logic Programs*, in "Journal of the ACM", January 2005, vol. 52, n^o 1, p. 102–146, http://prosecco.gforge.inria.fr/personal/bblanche/publications/AbadiBlanchetJACM7037.pdf

[45] M. ABADI, B. BLANCHET, C. FOURNET.*Just Fast Keying in the Pi Calculus*, in "ACM Transactions on Information and System Security (TISSEC)", July 2007, vol. 10, n^o 3, p. 1–59, http://prosecco.gforge.inria.fr/personal/bblanche/publications/AbadiBlanchetFournetTISSEC07.pdf

[46] C. ABATE, R. BLANCO, D. GARG, C. HRITCU, M. PATRIGNANI, J. THIBAULT.*Journey Beyond Full Abstraction: Exploring Robust Property Preservation for Secure Compilation*, July 2018, arXiv:1807.04603, https://arxiv.org/abs/1807.04603

[47] C. ABATE, A. AZEVEDO DE AMORIM, R. BLANCO, A. N. EVANS, G. FACHINI, C. HRITCU, T. LAURENT, B. C. PIERCE, M. STRONATI, A. TOLMACH.*When Good Components Go Bad: Formally Secure Compilation Despite Dynamic Compromise*, in "25th ACM Conference on Computer and Communications Security (CCS)", ACM, October 2018, p. 1351–1368, https://arxiv.org/abs/1802.00588

[48] D. AHMAN, C. FOURNET, C. HRITCU, K. MAILLARD, A. RASTOGI, N. SWAMY.*Recalling a Witness: Foundations and Applications of Monotonic State*, in "PACMPL", January 2018, vol. 2, n^o POPL, p. 65:1–65:30, https://arxiv.org/abs/1707.02466

[49] D. AHMAN, C. HRIŢCU, K. MAILLARD, G. MARTÍNEZ, G. PLOTKIN, J. PROTZENKO, A. RASTOGI, N. SWAMY.*Dijkstra Monads for Free*, in "44th ACM SIGPLAN Symposium on Principles of Programming Languages (POPL)", ACM, January 2017, p. 515-529 [*DOI :* 10.1145/3009837.3009878], https://www.fstar-lang.org/papers/dm4free/

[50] A. AZEVEDO DE AMORIM, M. DÉNÈS, N. GIANNARAKIS, C. HRIŢCU, B. C. PIERCE, A. SPECTOR-ZABUSKY, A. TOLMACH.*Micro-Policies: Formally Verified, Tag-Based Security Monitors*, in "36th IEEE Symposium on Security and Privacy (Oakland S&P)", IEEE Computer Society, May 2015, p. 813–830 [*DOI :* 10.1109/SP.2015.55], http://prosecco.gforge.inria.fr/personal/hritcu/publications/micro-policies.pdf

[51] A. AZEVEDO DE AMORIM, C. HRITCU, B. C. PIERCE.*The Meaning of Memory Safety*, in "7th International Conference on Principles of Security and Trust (POST)", April 2018, p. 79–105 [*DOI :* 10.1007/978-3-319-89722-6_4], https://arxiv.org/abs/1705.07354

[52] K. BHARGAVAN, B. BLANCHET, N. KOBEISSI.*Verified Models and Reference Implementations for the TLS 1.3 Standard Candidate*, in "38th IEEE Symposium on Security and Privacy", San Jose, United States, May 2017, p. 483 - 502 [*DOI :* 10.1109/SP.2017.26], https://hal.inria.fr/hal-01575920

[53] K. BHARGAVAN, B. BOND, A. DELIGNAT-LAVAUD, C. FOURNET, C. HAWBLITZEL, C. HRITCU, S. ISHTIAQ, M. KOHLWEISS, R. LEINO, J. LORCH, K. MAILLARD, J. PANG, B. PARNO, J. PROTZENKO, T. RAMANANANDRO, A. RANE, A. RASTOGI, N. SWAMY, L. THOMPSON, P. WANG, S. ZANELLA-BÉGUELIN, J.-K. ZINZINDOHOUÉ.*Everest: Towards a Verified, Drop-in Replacement of HTTPS*, in "2nd Summit on Advances in Programming Languages (SNAPL)", May 2017, http://drops.dagstuhl.de/opus/volltexte/2017/7119/pdf/LIPIcs-SNAPL-2017-1.pdf

[54] K. BHARGAVAN, A. DELIGNAT-LAVAUD, C. FOURNET, M. KOHLWEISS, J. PAN, J. PROTZENKO, A. RASTOGI, N. SWAMY, S. ZANELLA-BÉGUELIN, J.-K. ZINZINDOHOUÉ.*Implementing and Proving the TLS 1.3 Record Layer*, in "IEEE Symposium on Security and Privacy (Oakland)", 2017

[55] K. BHARGAVAN, C. FOURNET, R. CORIN, E. ZALINESCU.*Verified Cryptographic Implementations for TLS*, in "ACM Transactions Inf. Syst. Secur.", March 2012, vol. 15, n$^{o}$ 1, p. 3:1–3:32, http://doi.acm.org/10.1145/2133375.2133378

[56] K. BHARGAVAN, C. FOURNET, A. D. GORDON, N. SWAMY.*Verified implementations of the information card federated identity-management protocol*, in "ACM Symposium on Information, Computer and Communications Security (ASIACCS)", 2008, p. 123-135

[57] B. BLANCHET, M. ABADI, C. FOURNET.*Automated Verification of Selected Equivalences for Security Protocols*, in "Journal of Logic and Algebraic Programming", February–March 2008, vol. 75, n$^{o}$ 1, p. 3–51, http://prosecco.gforge.inria.fr/personal/bblanche/publications/BlanchetAbadiFournetJLAP07.pdf

[58] B. BLANCHET.*An Efficient Cryptographic Protocol Verifier Based on Prolog Rules*, in "14th IEEE Computer Security Foundations Workshop (CSFW'01)", 2001, p. 82–96

[59] B. BLANCHET.*Automatic Verification of Correspondences for Security Protocols*, in "Journal of Computer Security", July 2009, vol. 17, n$^{o}$ 4, p. 363–434, http://prosecco.gforge.inria.fr/personal/bblanche/publications/BlanchetJCS08.pdf

[60] B. BLANCHET, A. PODELSKI.*Verification of Cryptographic Protocols: Tagging Enforces Termination*, in "Theoretical Computer Science", March 2005, vol. 333, n$^{o}$ 1-2, p. 67–90, Special issue FoSSaCS'03, http://prosecco.gforge.inria.fr/personal/bblanche/publications/BlanchetPodelskiTCS04.html

[61] D. CADÉ, B. BLANCHET.*Proved Generation of Implementations from Computationally Secure Protocol Specifications*, in "Journal of Computer Security", 2015, vol. 23, n$^{o}$ 3, p. 331–402

[62] J. CLULOW.*On the Security of PKCS#11*, in "CHES", 2003, p. 411-425

[63] S. DELAUNE, S. KREMER, G. STEEL.*Formal Analysis of PKCS#11 and Proprietary Extensions*, in "Journal of Computer Security", November 2010, vol. 18, n° 6, p. 1211-1245 [*DOI :* 10.3233/JCS-2009-0394], http://www.lsv.ens-cachan.fr/Publis/PAPERS/PDF/DKS-jcs09.pdf

[64] A. DELIGNAT-LAVAUD, K. BHARGAVAN, S. MAFFEIS.*Language-Based Defenses Against Untrusted Browser Origins*, in "Proceedings of the 22th USENIX Security Symposium", 2013, http://prosecco.inria.fr/personal/karthik/pubs/language-based-defenses-against-untrusted-origins-sec13.pdf

[65] D. DOLEV, A. YAO.*On the security of public key protocols*, in "IEEE Transactions on Information Theory", 1983, vol. IT–29, n° 2, p. 198–208

[66] C. FOURNET, M. KOHLWEISS, P.-Y. STRUB.*Modular Code-Based Cryptographic Verification*, in "ACM Conference on Computer and Communications Security", 2011

[67] N. GRIMM, K. MAILLARD, C. FOURNET, C. HRITCU, M. MAFFEI, J. PROTZENKO, T. RAMANANANDRO, A. RASTOGI, N. SWAMY, S. ZANELLA-BÉGUELIN.*A Monadic Framework for Relational Verification: Applied to Information Security, Program Equivalence, and Optimizations*, in "7th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP)", ACM, January 2018, p. 130–145 [*DOI :* 10.1145/3167090], https://arxiv.org/abs/1703.00055

[68] N. KOBEISSI, K. BHARGAVAN, B. BLANCHET.*Automated Verification for Secure Messaging Protocols and Their Implementations: A Symbolic and Computational Approach*, in "2nd IEEE European Symposium on Security and Privacy", Paris, France, April 2017, p. 435 - 450 [*DOI :* 10.1109/EUROSP.2017.38], https://hal.inria.fr/hal-01575923

[69] G. MARTÍNEZ, D. AHMAN, V. DUMITRESCU, N. GIANNARAKIS, C. HAWBLITZEL, C. HRITCU, M. NARASIMHAMURTHY, Z. PARASKEVOPOULOU, C. PIT-CLAUDEL, J. PROTZENKO, T. RAMANANANDRO, A. RASTOGI, N. SWAMY.*Meta-F*: Proof Automation with SMT, Tactics, and Metaprograms*, March 2018, arXiv:1803.06547, https://arxiv.org/abs/1803.06547

[70] R. NEEDHAM, M. SCHROEDER.*Using encryption for authentication in large networks of computers*, in "Communications of the ACM", 1978, vol. 21, n° 12, p. 993–999

[71] J. PROTZENKO, J.-K. ZINZINDOHOUÉ, A. RASTOGI, T. RAMANANANDRO, P. WANG, S. ZANELLA-BÉGUELIN, A. DELIGNAT-LAVAUD, C. HRITCU, K. BHARGAVAN, C. FOURNET, N. SWAMY.*Verified Low-Level Programming Embedded in F**, in "PACMPL", September 2017, vol. 1, n° ICFP, p. 17:1–17:29 [*DOI :* 10.1145/3110261], http://arxiv.org/abs/1703.00053

[72] N. SWAMY, C. FOURNET, A. RASTOGI, K. BHARGAVAN, J. CHEN, P.-Y. STRUB, G. M. BIERMAN.*Gradual typing embedded securely in JavaScript*, in "41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)", 2014, p. 425-438, http://prosecco.inria.fr/personal/karthik/pubs/tsstar-popl14.pdf

[73] N. SWAMY, C. HRIŢCU, C. KELLER, A. RASTOGI, A. DELIGNAT-LAVAUD, S. FOREST, K. BHARGAVAN, C. FOURNET, P.-Y. STRUB, M. KOHLWEISS, J.-K. ZINZINDOHOUÉ, S. ZANELLA-BÉGUELIN.*Dependent Types and Multi-Monadic Effects in F**, in "43rd ACM SIGPLAN-SIGACT Symposium on Principles

of Programming Languages (POPL)", ACM, January 2016, p. 256-270, https://www.fstar-lang.org/papers/mumon/

[74] J.-K. ZINZINDOHOUÉ, K. BHARGAVAN, J. PROTZENKO, B. BEURDOUCHE.*HACL*: A Verified Modern Cryptographic Library*, in "Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017", 2017, p. 1789–1806, http://doi.acm.org/10.1145/3133956.3134043

# Project-Team QUANTIC

## QUANTum Information Circuits

IN COLLABORATION WITH: Centre Automatique et Systèmes, Laboratoire Pierre Aigrain

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

**Mines ParisTech**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Optimization and control of dynamic systems**

# Table of contents

# Project-Team QUANTIC

*Creation of the Team: 2013 September 12, updated into Project-Team: 2015 April 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.1.11. - Quantum architectures
A4.2. - Correcting codes
A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.2. - Stochastic Modeling
A6.1.3. - Discrete Modeling (multi-agent, people centered)
A6.1.4. - Multiscale modeling
A6.2. - Scientific computing, Numerical Analysis & Optimization
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.3. - Probabilistic methods
A6.2.6. - Optimization
A6.3.1. - Inverse problems
A6.3.2. - Data assimilation
A6.3.3. - Data processing
A6.3.4. - Model reduction
A6.4. - Automatic control
A6.4.1. - Deterministic control
A6.4.2. - Stochastic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization

**Other Research Topics and Application Domains:**

B5.3. - Nanotechnology
B5.4. - Microelectronics
B6.5. - Information systems
B9.10. - Privacy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Mazyar Mirrahimi [Team leader, Inria, Senior Researcher]
Alain Sarlette [Inria, Researcher]

**Faculty Members**

Zaki Leghtas [Ecole Nationale Supérieure des Mines de Paris , Associate Professor]
Pierre Rouchon [Ecole Nationale Supérieure des Mines de Paris , Professor, HDR]

**External Collaborator**

Michel Sorine [Inria, HDR]

**PhD Students**

Michiel Burgelman [Inria, from Oct 2018]
Gerardo Cardona Sanchez [Ecole Nationale Supérieure des Mines de Paris]
Jeremie Guillaud [Inria]
Vincent Martin [Inria, from Oct 2018]
Lucas Verney [Ecole Normale Supérieure Paris]

**Post-Doctoral Fellows**

Paolo Forni [Ecole Nationale Supérieure des Mines de Paris]
Zibo Miao [Inria, until Feb 2018]

**Visiting Scientist**

Tryphon Georgiou [University of California at Irvine, from Apr 2018 until Jul 2018]

**Administrative Assistants**

Derya Gok [Inria, from Sep 2018]
Martine Verneuille [Inria, until Sep 2018]

# 2. Overall Objectives

## 2.1. Overall objectives

The research activities of QUANTIC team lie at the border between theoretical and experimental efforts in the emerging field of quantum systems engineering. Our research topics are in direct continuation of a historic research theme of Inria, classical automatic control, while opening completely new perspectives toward quantum control: by developing a new mathematical system theory for quantum circuits, we will realize the components of a future quantum information processing unit.

One of the unique features of our team concerns the large spectrum of our subjects going from the mathematical analysis of the physical systems (development of systematic mathematical methods for control and estimation of quantum systems), and the numerical analysis of the proposed solutions, to the experimental implementation of the quantum circuits based on these solutions. This is made possible by the constant and profound interaction between the applied mathematicians and the physicists in the group. Indeed, this close collaboration has already brought a significant acceleration in our research efforts. In a long run, this synergy should lead to a deeper understanding of the physical phenomena behind these emerging technologies and the development of new research directions within the field of quantum information processing.

Towards this ultimate task of practical quantum digital systems, the approach of the QUANTIC team is complementary to the one taken by teams with expertise in quantum algorithms. Indeed, we start from the specific controls that can be realistically applied on physical systems, to propose designs which combine them into *hardware shortcuts* implementing *robust* behaviors useful for quantum information processing. Whenever a significant new element of quantum engineering architecture is developed, the initial motivation is to prove an enabling technology with major impact for the groups working one abstraction layer higher: on quantum algorithms but also on e.g. secure communication and metrology applications.

# 3. Research Program

## 3.1. Hardware-efficient quantum information processing

In this scientific program, we will explore various theoretical and experimental issues concerning protection and manipulation of quantum information. Indeed, the next, critical stage in the development of Quantum Information Processing (QIP) is most certainly the active quantum error correction (QEC). Through this stage one designs, possibly using many physical qubits, an encoded logical qubit which is protected against major decoherence channels and hence admits a significantly longer effective coherence time than a physical qubit. Reliable (fault-tolerant) computation with protected logical qubits usually comes at the expense of a significant

overhead in the hardware (up to thousands of physical qubits per logical qubit). Each of the involved physical qubits still needs to satisfy the best achievable properties (coherence times, coupling strengths and tunability). More remarkably, one needs to avoid undesired interactions between various subsystems. This is going to be a major difficulty for qubits on a single chip.

The usual approach for the realization of QEC is to use many qubits to obtain a larger Hilbert space of the qubit register [85], [88]. By redundantly encoding quantum information in this Hilbert space of larger dimension one make the QEC tractable: different error channels lead to distinguishable error syndromes. There are two major drawbacks in using multi-qubit registers. The first, fundamental, drawback is that with each added physical qubit, several new decoherence channels are added. Because of the exponential increase of the Hilbert's space dimension versus the linear increase in the number of decay channels, using enough qubits, one is able to eventually protect quantum information against decoherence. However, multiplying the number of possible errors, this requires measuring more error syndromes. Note furthermore that, in general, some of these new decoherence channels can lead to correlated action on many qubits and this needs to be taken into account with extra care: in particular, such kind of non-local error channels are problematic for surface codes. The second, more practical, drawback is that it is still extremely challenging to build a register of more than on the order of 10 qubits where each of the qubits is required to satisfy near the best achieved properties: these properties include the coherence time, the coupling strengths and the tunability. Indeed, building such a register is not merely only a fabrication task but rather, one requirers to look for architectures such that, each individual qubit can be addressed and controlled independently from the others. One is also required to make sure that all the noise channels are well-controlled and uncorrelated for the QEC to be effective.

We have recently introduced a new paradigm for encoding and protecting quantum information in a quantum harmonic oscillator (e.g. a high-Q mode of a 3D superconducting cavity) instead of a multi-qubit register [62]. The infinite dimensional Hilbert space of such a system can be used to redundantly encode quantum information. The power of this idea lies in the fact that the dominant decoherence channel in a cavity is photon damping, and no more decay channels are added if we increase the number of photons we insert in the cavity. Hence, only a single error syndrome needs to be measured to identify if an error has occurred or not. Indeed, we are convinced that most early proposals on continuous variable QIP [59], [53] could be revisited taking into account the design flexibilities of Quantum Superconducting Circuits (QSC) and the new coupling regimes that are provided by these systems. In particular, we have illustrated that coupling a qubit to the cavity mode in the strong dispersive regime provides an important controllability over the Hilbert space of the cavity mode [61]. Through a recent experimental work [93], we benefit from this controllability to prepare superpositions of quasi-orthogonal coherent states, also known as Schrödinger cat states.

In this Scheme, the logical qubit is encoded in a four-component Schrödinger cat state. Continuous quantum non-demolition (QND) monitoring of a single physical observable, consisting of photon number parity, enables then the tractability of single photon jumps. We obtain therefore a first-order quantum error correcting code using only a single high-Q cavity mode (for the storage of quantum information), a single qubit (providing the non-linearity needed for controllability) and a single low-Q cavity mode (for reading out the error syndrome). An earlier experiment on such QND photon-number parity measurements [89] has recently led to a first experimental realization of a full quantum error correcting code improving the coherence time of quantum information [5]. As shown in Figure 1, this leads to a significant hardware economy for realization of a protected logical qubit. Our goal here is to push these ideas towards a reliable and hardware-efficient paradigm for universal quantum computation.

## 3.2. Reservoir (dissipation) engineering and autonomous stabilization of quantum systems

Being at the heart of any QEC protocol, the concept of feedback is central for the protection of quantum information, enabling many-qubit quantum computation or long-distance quantum communication. However, such a closed-loop control which requires a real-time and continuous measurement of the quantum system has been for long considered as counter-intuitive or even impossible. This thought was mainly caused by properties of quantum measurements: any measurement implies an instantaneous strong perturbation to the

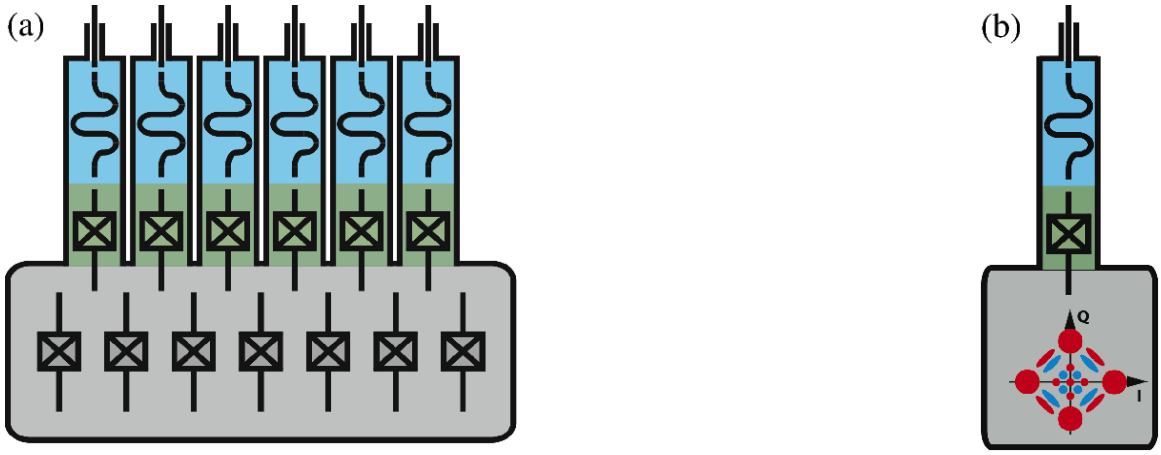


*Figure 1. (a) A protected logical qubit consisting of a register of many qubits: here, we see a possible architecture for the Steane code [88] consisting of 7 qubits requiring the measurement of 6 error syndromes. In this sketch, 7 transmon qubits in a high-Q resonator and the measurement of the 6 error syndromes is ensured through 6 additional ancillary qubits with the possibility of individual readout of the ancillary qubits via independent low-Q resonators. (b) Minimal architecture for a protected logical qubit, adapted to circuit quantum electrodynamics experiments. Quantum information is encoded in a Schrödinger cat state of a single high-Q resonator mode and a single error syndrome is measured, using a single ancillary transmon qubit and the associated readout low-Q resonator.*

system's state. The concept of *quantum non-demolition* (QND) measurement has played a crucial role in understanding and resolving this difficulty [36]. In the context of cavity quantum electro-dynamics (cavity QED) with Rydberg atoms [55], a first experiment on continuous QND measurements of the number of microwave photons was performed by the group at Laboratoire Kastler-Brossel (ENS) [54]. Later on, this ability of performing continuous measurements allowed the same group to realize the first continuous quantum feedback protocol stabilizing highly non-classical states of the microwave field in the cavity, the so-called photon number states [8] (this ground-breaking work was mentioned in the Nobel prize attributed to Serge Haroche). The QUANTIC team contributed to the theoretical work behind this experiment [45], [27], [87], [29]. These contributions include the development and optimization of the quantum filters taking into account the quantum measurement back-action and various measurement noises and uncertainties, the development of a feedback law based on control Lyapunov techniques, and the compensation of the feedback delay.

In the context of circuit quantum electrodynamics (circuit QED) [44], recent advances in quantum-limited amplifiers [79], [91] have opened doors to high-fidelity non-demolition measurements and real-time feedback for superconducting qubits [56]. This ability to perform high-fidelity non-demolition measurements of a quantum signal has very recently led to quantum feedback experiments with quantum superconducting circuits [91], [78], [38]. Here again, the QUANTIC team has participated to one of the first experiments in the field where the control objective is to track a dynamical trajectory of a single qubit rather than stabilizing a stationary state. Such quantum trajectory tracking could be further explored to achieve metrological goals such as the stabilization of the amplitude of a microwave drive [69].

While all this progress has led to a strong optimism about the possibility to perform active protection of quantum information against decoherence, the rather short dynamical time scales of these systems limit, to a great amount, the complexity of the feedback strategies that could be employed. Indeed, in such measurement-

based feedback protocols, the time-consuming data acquisition and post-treatment of the output signal leads to an important latency in the feedback procedure.

The reservoir (dissipation) engineering [76] and the closely related coherent feedback [67] are considered as alternative approaches circumventing the necessity of a real-time data acquisition, signal processing and feedback calculations. In the context of quantum information, the decoherence, caused by the coupling of a system to uncontrolled external degrees of freedom, is generally considered as the main obstacle to synthesize quantum states and to observe quantum effects. Paradoxically, it is possible to intentionally engineer a particular coupling to a reservoir in the aim of maintaining the coherence of some particular quantum states. In a general viewpoint, these approaches could be understood in the following manner: by coupling the quantum system to be stabilized to a strongly dissipative ancillary quantum system, one evacuates the entropy of the main system through the dissipation of the ancillary one. By building the feedback loop into the Hamiltonian, this type of autonomous feedback obviates the need for a complicated external control loop to correct errors. On the experimental side, such autonomous feedback techniques have been used for qubit reset [52], single-qubit state stabilization [71], and the creation [31] and stabilization [60], [66][9] of states of multipartite quantum systems.

Such reservoir engineering techniques could be widely revisited exploring the flexibility in the Hamiltonian design for QSC. We have recently developed theoretical proposals leading to extremely efficient, and simple to implement, stabilization schemes for systems consisting of a single, two or three qubits [52], [64], [42][12]. The experimental results based on these protocols have illustrated the efficiency of the approach [52][9]. Through these experiments, we exploit the strong dispersive interaction [83] between superconducting qubits and a single low-Q cavity mode playing the role of a dissipative reservoir. Applying continuous-wave (cw) microwave drives with well-chosen fixed frequencies, amplitudes, and phases, we engineer an effective interaction Hamiltonian which evacuates the entropy of the system interacting with a noisy environment: by driving the qubits and cavity with continuous-wave drives, we induce an autonomous feedback loop which corrects the state of the qubits every time it decays out of the desired target state. The schemes are robust against small variations of the control parameters (drives amplitudes and phase) and require only some basic calibration. Finally, by avoiding resonant interactions between the qubits and the low-Q cavity mode, the qubits remain protected against the Purcell effect, which would reduce the coherence times. We have also investigated both theoretically and experimentally the autonomous stabilization of non-classical states (such as Schrodinger cat states and Fock states) of microwave field confined in a high-Q cavity mode [70], [81], [57][4].

## 3.3. System theory for quantum information processing

In parallel and in strong interactions with the above experimental goals, we develop systematic mathematical methods for dynamical analysis, control and estimation of composite and open quantum systems. These systems are built with several quantum subsystems whose irreversible dynamics results from measurements and/or decoherence. A special attention is given to spin/spring systems made with qubits and harmonic oscillators. These developments are done in the spirit of our recent contributions [80], [27], [86], [82], [87], [29][7] resulting from collaborations with the cavity quantum electrodynamics group of Laboratoire Kastler Brossel.

### *3.3.1. Stabilization by measurement-based feedback*

The protection of quantum information via efficient QEC is a combination of (i) tailored dynamics of a quantum system in order to protect an informational qubit from certain decoherence channels, and (ii) controlled reaction to measurements that efficiently detect and correct the dominating disturbances that are not rejected by the tailored quantum dynamics.

In such feedback scheme, the system and its measurement are quantum objects whereas the controller and the control input are classical. The stabilizing control law is based on the past values of the measurement outcomes. During our work on the LKB photon box, we have developed, for single input systems subject to quantum non-demolition measurement, a systematic stabilization method [29]: it is based on a discrete-time

formulation of the dynamics, on the construction of a strict control Lyapunov function and on an explicit compensation of the feedback-loop delay. Keeping the QND measurement assumptions, extensions of such stabilization schemes will be investigated in the following directions: finite set of values for the control input with application to the convergence analysis of the atomic feedback scheme experimentally tested in [94]; multi-input case where the construction by inversion of a Metzler matrix of the strict Lyapunov function is not straightforward; continuous-time systems governed by diffusive master equations; stabilization towards a set of density operators included in a target subspace; adaptive measurement by feedback to accelerate the convergence towards a stationary state as experimentally tested in [74]. Without the QND measurement assumptions, we will also address the stabilization of non-stationary states and trajectory tracking, with applications to systems similar to those considered in [56], [38].

### 3.3.2. Filtering, quantum state and parameter estimations

The performance of every feedback controller crucially depends on its online estimation of the current situation. This becomes even more important for quantum systems, where full state measurements are physically impossible. Therefore the ultimate performance of feedback correction depends on fast, efficient and optimally accurate state and parameter estimations.

A quantum filter takes into account imperfection and decoherence and provides the quantum state at time $t \geq 0$ from an initial value at $t = 0$ and the measurement outcomes between 0 and $t$. Quantum filtering goes back to the work of Belavkin [32] and is related to quantum trajectories [40], [43]. A modern and mathematical exposure of the diffusive models is given in [30]. In [95] a first convergence analysis of diffusive filters is proposed. Nevertheless the convergence characterization and estimation of convergence rate remain open and difficult problems. For discrete time filters, a general stability result based on fidelity is proven in [80], [86]. This stability result is extended to a large class of continuous-time filters in [28]. Further efforts are required to characterize asymptotic and exponential stability. Estimations of convergence rates are available only for quantum non-demolition measurements [33]. Parameter estimations based on measurement data of quantum trajectories can be formulated within such quantum filtering framework [47], [72].

We will continue to investigate stability and convergence of quantum filtering. We will also exploit our fidelity-based stability result to justify maximum likelihood estimation and to propose, for open quantum system, parameter estimation algorithms inspired of existing estimation algorithms for classical systems. We will also investigate a more specific quantum approach: it is noticed in [37] that post-selection statistics and "past quantum" state analysis [48] enhance sensitivity to parameters and could be interesting towards increasing the precision of an estimation.

### 3.3.3. Stabilization by interconnections

In such stabilization schemes, the controller is also a quantum object: it is coupled to the system of interest and is subject to decoherence and thus admits an irreversible evolution. These stabilization schemes are closely related to reservoir engineering and coherent feedback [76], [67]. The closed-loop system is then a composite system built with the original system and its controller. In fact, and given our particular recent expertise in this domain [7], [9] [52], this subsection is dedicated to further developing such stabilization techniques, both experimentally and theoretically.

The main analysis issues are to prove the closed-loop convergence and to estimate the convergence rates. Since these systems are governed by Lindblad differential equations (continuous-time case) or Kraus maps (discrete-time case), their stability is automatically guaranteed: such dynamics are contractions for a large set of metrics (see [75]). Convergence and asymptotic stability is less well understood. In particular most of the convergence results consider the case where the target steady-state is a density operator of maximum rank (see, e.g., [26][chapter 4, section 6]). When the goal steady-state is not full rank very few convergence results are available.

We will focus on this geometric situation where the goal steady-state is on the boundary of the cone of positive Hermitian operators of finite trace. A specific attention will be given to adapt standard tools (Lyapunov function, passivity, contraction and Lasalle's invariance principle) for infinite dimensional systems

to spin/spring structures inspired of [7], [9] [52], [70] and their associated Fokker-Planck equations for the Wigner functions.

We will also explore the Heisenberg point of view in connection with recent results of the Inria project-team MAXPLUS (algorithms and applications of algebras of max-plus type) relative to Perron-Frobenius theory [51], [50]. We will start with [84] and [77] where, based on a theorem due to Birkhoff [34], dual Lindblad equations and dual Kraus maps governing the Heisenberg evolution of any operator are shown to be contractions on the cone of Hermitian operators equipped with Hilbert's projective metric. As the Heisenberg picture is characterized by convergence of all operators to a multiple of the identity, it might provide a mean to circumvent the rank issues. We hope that such contraction tools will be especially well adapted to analyzing quantum systems composed of multiple components, motivated by the facts that the same geometry describes the contraction of classical systems undergoing synchronizing interactions [90] and by our recent generalized extension of the latter synchronizing interactions to quantum systems [68].

Besides these analysis tasks, the major challenge in stabilization by interconnections is to provide systematic methods for the design, from typical building blocks, of control systems that stabilize a specific quantum goal (state, set of states, operation) when coupled to the target system. While constructions exist for so-called linear quantum systems [73], this does not cover the states that are more interesting for quantum applications. Various strategies have been proposed that concatenate iterative control steps for open-loop steering [92], [65] with experimental limitations. The characterization of Kraus maps to stabilize any types of states has also been established [35], but without considering experimental implementations. A viable stabilization by interaction has to combine the capabilities of these various approaches, and this is a missing piece that we want to address.

*3.3.3.1. Perturbation methods*

With this subsection we turn towards more fundamental developments that are necessary in order to address the complexity of quantum networks with efficient reduction techniques. This should yield both efficient mathematical methods, as well as insights towards unravelling dominant physical phenomena/mechanisms in multipartite quantum dynamical systems.

In the Schrödinger point of view, the dynamics of open quantum systems are governed by master equations, either deterministic or stochastic [55], [49]. Dynamical models of composite systems are based on tensor products of Hilbert spaces and operators attached to the constitutive subsystems. Generally, a hierarchy of different timescales is present. Perturbation techniques can be very useful to construct reliable models adapted to the timescale of interest.

To eliminate high frequency oscillations possibly induced by quasi-resonant classical drives, averaging techniques are used (rotating wave approximation). These techniques are well established for closed systems without any dissipation nor irreversible effect due to measurement or decoherence. We will consider in a first step the adaptation of these averaging techniques to deterministic Lindblad master equations governing the quantum state, i.e. the system density operator. Emphasis will be put on first order and higher order corrections based on non-commutative computations with the different operators appearing in the Lindblad equations. Higher order terms could be of some interest for the protected logical qubit of figure 1b. In future steps, we intend to explore the possibility to explicitly exploit averaging or singular perturbation properties in the design of coherent quantum feedback systems; this should be an open-systems counterpart of works like [63].

To eliminate subsystems subject to fast convergence induced by decoherence, singular perturbation techniques can be used. They provide reduced models of smaller dimension via the adiabatic elimination of the rapidly converging subsystems. The derivation of the slow dynamics is far from being obvious (see, e.g., the computations of page 142 in [39] for the adiabatic elimination of low-Q cavity). Conversely to the classical composite systems where we have to eliminate one component in a Cartesian product, we here have to eliminate one component in a tensor product. We will adapt geometric singular perturbations [46] and invariant manifold techniques [41] to such tensor product computations to derive reduced slow approximations of any order. Such adaptations will be very useful in the context of quantum Zeno dynamics to obtain approximations of the slow dynamics on the decoherence-free subspace corresponding to the slow attractive manifold.

Perturbation methods are also precious to analyze convergence rates. Deriving the spectrum attached to the Lindblad differential equation is not obvious. We will focus on the situation where the decoherence terms of the form $L\rho L^\dagger - (L^\dagger L\rho + \rho L^\dagger L)/2$ are small compared to the conservative terms $-i[H/\hbar, \rho]$. The difficulty to overcome here is the degeneracy of the unperturbed spectrum attached to the conservative evolution $\frac{d}{dt}\rho = -i[H/\hbar, \rho]$. The degree of degeneracy of the zero eigenvalue always exceeds the dimension of the Hilbert space. Adaptations of usual perturbation techniques [58] will be investigated. They will provide estimates of convergence rates for slightly open quantum systems. We expect that such estimates will help to understand the dependence on the experimental parameters of the convergence rates observed in [52][9] [64].

As particular outcomes for the other subsections, we expect that these developments towards simpler dominant dynamics will guide the search for optimal control strategies, both in open-loop microwave networks and in autonomous stabilization schemes such as reservoir engineering. It will further help to efficiently compute explicit convergence rates and quantitative performances for all the intended experiments.

# 4. Application Domains

## 4.1. Quantum engineering

A new field of quantum systems engineering has emerged during the last few decades. This field englobes a wide range of applications including nano-electromechanical devices, nuclear magnetic resonance applications, quantum chemical synthesis, high resolution measurement devices and finally quantum information processing devices for implementing quantum computation and quantum communication. Recent theoretical and experimental achievements have shown that the quantum dynamics can be studied within the framework of estimation and control theory, but give rise to new models that have not been fully explored yet.

The QUANTIC team's activities are defined at the border between theoretical and experimental efforts of this emerging field with an emphasis on the applications in quantum information, computation and communication. The main objective of this interdisciplinary team is to develop quantum devices ensuring a robust processing of quantum information.

On the theory side, this is done by following a system theory approach: we develop estimation and control tools adapted to particular features of quantum systems. The most important features, requiring the development of new engineering methods, are related to the concept of measurement and feedback for composite quantum systems. The destructive and partial [0] nature of measurements for quantum systems lead to major difficulties in extending classical control theory tools. Indeed, design of appropriate measurement protocols and, in the sequel, the corresponding quantum filters estimating the state of the system from the partial measurement record, are themselves building blocks of the quantum system theory to be developed.

On the experimental side, we develop new quantum information processing devices based on quantum superconducting circuits. Indeed, by realizing superconducting circuits at low temperatures and using microwave measurement techniques, the macroscopic and collective degrees of freedom such as the voltage and the current are forced to behave according to the laws of quantum mechanics. Our quantum devices are aimed to protect and process quantum information through these integrated circuits.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

[0] Here the partiality means that no single quantum measurement is capable of providing the complete information on the state of the system.

- Pierre Rouchon was the main organizer of the spring thematic quarter at Institut Henri Poincaré entitled "Measurement and control of quantum systems: theory and experiments" (16 April – 13 July 2018). This thematic quarter included courses, lectures and conferences. In particular, a research school of one week at CIRM, two 3-day workshops in May and June and the 2018 issue of PRACQSYS conference in July were organized throughout the quarter. This thematic quarter involved several hundred of participants. See IHP web page (http://www.ihp.fr/en/CEB/T2-2018), CIRM web page (https://conferences.cirm-math.fr/1732.html) and the specific quarter web site (https://sites.google.com/view/mcqs2018/home).
- QUANTIC has received a sub-award from Yale university for pursuing the collaborations of Mazyar Mirrahimi and his students/postdocs. In the framework of a new ARO (Army Research Office) grant received by our collaborators at Yale, QUANTIC receives 500k dollars over 4 years to fund the hiring of PhD students/ postdocs working on the collaborative subjects with Yale and also to cover the travels between Inria and Yale.
- Alain Sarlette has received a JCJC ANR grant entitled HAMROQS "High-accuracy model reduction for open quantum systems". This grant of 212k euros over 4 years will fund the activities of Alain Sarlette and his students/postdocs on systematic methods for quantum systems model reduction.
- PhD students of Alain Sarlette, Arash Farnam and Simon Apers, defended their PhD at his previous institution (Ghent university, Belgium).
- Mazyar Mirrahimi was an invited speaker at the American Physical Society March Meeting in Los Angeles.
- Mazyar Mirrahimi was a semi-plenary speaker at MTNS in Hong Kong (Mathematical Theory of Networks and Systems).

# 6. New Results

## 6.1. Simulation of quantum walks and fast mixing with classical processes

Participants: A. Sarlette

This is the final result of a line of work where we show that the mixing behavior of quantum walks on graphs can always be simulated by a classical "lifted Markov chain". This implies that quantum walks must satisfy a conductance bound on mixing speed, like classical Markov chains. Also current efficient quantum walk constructions are linked to classical processes that provide the same convergence speed. This excludes a simple characterization of quantum walk advantages in terms of bare mixing speed, as has been done by some previous authors comparing just to simple Markov chains. The question of efficient design of walks on graphs, on the basis of local graph queries and for specific applications, is thus brought back to the center of the focus for quantum walks. This collaborative work with F. Ticozzi (U. of Padova) has been published in [11].

As a follow-up on this work, we have developed algorithms in the latter sense: quantum walks on the basis of local design and which do speed up some applications. These last results have been presented as posters at conferences and will hopefully be part of next year's publications.

## 6.2. Adiabatic elimination for multi-partite open quantum systems with non-trivial zero-order dynamics

Participants: Paolo Forni, Alain Sarlette, Pierre Rouchon

We pursue the work initiated in our group during the thesis of Rémi Azouit, where we apply center manifold theory in order to reduce the model of a quantum system to its slowly contracting dynamics. Such model reduction is ubiquitous in models of coupled quantum systems where part of the system relaxes quickly towards an equilibrium situation, and acts as an environment for a system of interest. The extension presented in this work is the answer to a question by experimental physicists at Laboratoire Kastler Brossel (LKB), where they apply a strong drive which, in an 'intuitive model', would saturate so-called two-level-system impurities and thereby imply a particular behavior of frequency shift and dissipation on the target system (slow dynamics) as a function of drive characteristics. A good model for this situation involves, beyond a strongly dissipative environment, also a fast non-dissipative dynamics on the slowly contracting subsystem. Adding the latter into the model reduction was the purpose of this result. We analyze the experimental results and show that the model reduction allows us to explain the observed trends. This result led to a publication in collaboration with physicists Thibault Capelle, Emmanuel Flurin and Samule Deleglise from LKB [20].

Further extensions of adiabatic elimination formulas have been worked out during this year and will hopefully be part of next year's publications.

## 6.3. Exponential stochastic stabilization of a two-level quantum system via strict Lyapunov control

Participants: Gerardo Cardona, Alain Sarlette, Pierre Rouchon

In this result, we address the fundamental task of stabilizing the state of a quantum system towards a target eigenstate of a continuous-time quantum nondemolition measurement. The starting point is that a static output feedback does not allow us to stabilize this system, while more complicated procedures were not able to provide a convergence rate. Our main idea is to introduce a dynamic feedback controller of moderate complexity, where (i) feedback gains depend on estimated state and progressively go to zero as one approaches the target; and (ii) the feedback involves noise (in this paper from the measurement back-action but in further extensions possibly just independent noise). With this controller we show, providing a Lyapunov function close to the Bures distance measure, that the system converges exponentially towards the target eigenstate. This result, restricted to a proof-of-principle on the qubit, was published in [19].

This has laid the basis for further work, presented on posters and to be published next year, where we have shown that:

- the optimal convergence rate, equal to information gain, can be achieved with this feedback;

- the procedure extends to N-level systems, with noise just independent instead of coming from the measurement backaction;

- the procedure can be exploited towards continuous-time measurement-based quantum error correction

## 6.4. Structural instability of driven Josephson circuits prevented by an inductive shunt

Participants: Lucas Verney, Raphaël Lescanne, Zaki Leghtas, Mazyar Mirrahimi.

Superconducting circuits are a versatile platform to implement a multitude of Hamiltonians which perform quantum computation, simulation and sensing tasks. A key ingredient for realizing a desired Hamiltonian is the irradiation of the circuit by a strong drive. These strong drives provide an insitu control of couplings, which cannot be obtained by near-equilibrium Hamiltonians. However, as shown in our result, out-of-equilibrium systems are easily plagued by complex dynamics leading to instabilities. Predicting and preventing these instabilities is crucial, both from a fundamental and application perspective. We propose an inductively shunted transmon as the elementary circuit optimized for strong parametric drives. Developing a novel numerical approach that avoids the built-in limitations of perturbative analysis, we demonstrate that adding the inductive shunt significantly extends the range of pump powers over which the circuit behaves in a stable manner. This collaborative work between the Quantic team and Michel Devoret at Yale has been recently submitted for publication [25].

## 6.5. Observing the escape of a driven quantum Josephson circuit into unconfined states

Participants: Raphaël Lescanne, Lucas Verney, Mazyar Mirrahimi, Zaki Leghtas.

Josephson circuits have been ideal systems to study complex non-linear dynamics which can lead to chaotic behavior and instabilities. More recently, Josephson circuits in the quantum regime, particularly in the presence of microwave drives, have demonstrated their ability to emulate a variety of Hamiltonians that are useful for the processing of quantum information. In this experimental work, we show that these drives lead to an instability which results in the escape of the circuit mode into states that are not confined by the Josephson cosine potential. We observe this escape in a ubiquitous circuit: a transmon embedded in a 3D cavity. When the transmon occupies these free-particle-like states, the circuit behaves as though the junction had been removed, and all non-linearities are lost. This work deepens our understanding of strongly driven Josephson circuits, which is important for fundamental and application perspectives, such as the engineering of Hamiltonians by parametric pumping. This collaborative work between Quantic team, Benjamin Huard's team at ENS Lyon and Michel Devoret at Yale, has been recently submitted for publication [21].

## 6.6. Dynamics of a qubit while simultaneously monitoring its relaxation and dephasing

Participants: Zaki Leghtas.

Decoherence originates from the leakage of quantum information into external degrees of freedom. For a qubit, the two main decoherence channels are relaxation and dephasing. Here, we report an experiment on a superconducting qubit where we retrieve part of the lost information in both of these channels. We demonstrate that raw averaging of the corresponding measurement records provides a full quantum tomography of the qubit state where all three components of the effective spin-1/2 are simultaneously measured. From single realizations of the experiment, it is possible to infer the quantum trajectories followed by the qubit state conditioned on relaxation and/or dephasing channels. The incompatibility between these quantum measurements of the qubit leads to observable consequences in the statistics of quantum states. The high level of controllability of superconducting circuits enables us to explore many regimes from the Zeno effect to underdamped Rabi oscillations depending on the relative strengths of driving, dephasing, and relaxation. This work is a collaboration between the Quantic team and the group of Benjamin Huard at ENS Lyon and was published in [13].

## 6.7. Demonstration of an effective ultrastrong coupling between two oscillators

Participants: Zaki Leghtas

When the coupling rate between two quantum systems becomes as large as their characteristic frequencies, it induces dramatic effects on their dynamics and even on the nature of their ground state. The case of a qubit coupled to a harmonic oscillator in this ultrastrong coupling regime has been investigated theoretically and experimentally. Here, we explore the case of two harmonic oscillators in the ultrastrong coupling regime. Probing the properties of their ground state remains out of reach in natural implementations. Therefore, we have realized an analog quantum simulation of this coupled system by dual frequency pumping a nonlinear superconducting circuit. The pump amplitudes directly tune the effective coupling rate. We observe spectroscopic signature of a mode hybridization that is characteristic of the ultrastrong coupling. We experimentally demonstrate a key property of the ground state of this simulated ultrastrong coupling between modes by observing simultaneous single- and two-mode squeezing of the radiated field below vacuum fluctuations. This work is a collaboration between the Quantic team and the group of Benjamin Huard at ENS Lyon and was published in [14].

## 6.8. Fault-tolerant detection of a quantum error

Participants: Mazyar Mirrahimi

A critical component of any quantum error–correcting scheme is detection of errors by using an ancilla system. However, errors occurring in the ancilla can propagate onto the logical qubit, irreversibly corrupting the encoded information. We experimentally demonstrate a fault-tolerant error-detection scheme that suppresses spreading of ancilla errors by a factor of 5, while maintaining the assignment fidelity. The same method is used to prevent propagation of ancilla excitations, increasing the logical qubit dephasing time by an order of magnitude. Our approach is hardware-efficient, as it uses a single multilevel transmon ancilla and a cavity-encoded logical qubit, whose interaction is engineered in situ by using an off-resonant sideband drive. The results demonstrate that hardware-efficient approaches that exploit system-specific error models can yield advances toward fault-tolerant quantum computation. This work is a collaboration between the Quantic team and the group of Robert Schoelkopf at Yale university and was published in [17].

## 6.9. Coherent oscillations inside a quantum manifold stabilized by dissipation

Participants: Zaki Leghtas, Mazyar Mirrahimi

Manipulating the state of a logical quantum bit usually comes at the expense of exposing it to decoherence. Fault-tolerant quantum computing tackles this problem by manipulating quantum information within a stable manifold of a larger Hilbert space, whose symmetries restrict the number of independent errors. The remaining errors do not affect the quantum computation and are correctable after the fact. Here we implement the autonomous stabilization of an encoding manifold spanned by Schrödinger cat states in a superconducting cavity. We show Zeno-driven coherent oscillations between these states analogous to the Rabi rotation of a qubit protected against phase flips. Such gates are compatible with quantum error correction and hence are crucial for fault-tolerant logical qubits. This experimental work follows our previous theoretical proposal [70]. It is a collaboration between the Quantic team and the group of Michel Devoret at Yale university and was published in [18].

## 6.10. To catch and reverse a quantum jump mid-flight

Participants: Mazyar Mirrahimi

A quantum system driven by a weak deterministic force while under strong continuous energy measurement exhibits quantum jumps between its energy levels. This celebrated phenomenon is emblematic of the special nature of randomness in quantum physics. The times at which the jumps occur are reputed to be fundamentally unpredictable. However, certain classical phenomena, like tsunamis, while unpredictable in the long term, may possess a degree of predictability in the short term, and in some cases it may be possible to prevent a disaster by detecting an advance warning signal. Can there be, despite the indeterminism of quantum physics, a possibility to know if a quantum jump is about to occur or not? We answer this question affirmatively by experimentally demonstrating that the completed jump from the ground to an excited state of a superconducting artificial atom can be tracked, as it follows its predictable "flight," by monitoring the population of an auxiliary level coupled to the ground state. Furthermore, we show that the completed jump is continuous, deterministic, and coherent. Exploiting this coherence, we catch and reverse a quantum jump mid-flight, thus preventing its completion. This real-time intervention is based on a particular lull period in the population of the auxiliary level, which serves as our advance warning signal. Our experimental results, which agree with theoretical predictions essentially without adjustable parameters, support the modern quantum trajectory theory and provide new ground for the exploration of real-time intervention techniques in the control of quantum systems, such as early detection of error syndromes. This work is a collaboration between the Quantic team and the group of Michel Devoret at Yale university and is recently submitted for publication [22].

## 6.11. Remote entanglement stabilization and concentration by quantum reservoir engineering

Participants: Nicolas Didier, Jérémie Guillaud, Mazyar Mirrahimi

Quantum information processing in a modular architecture requires the distribution, stabilization, and distillation of entanglement in a qubit network. We present autonomous entanglement stabilization protocols between two superconducting qubits that are coupled to distant cavities. The coupling between cavities is mediated and controlled via a three-wave mixing device that generates either a two-mode squeezed state or a delocalized mode between the remote cavities depending on the pump applied to the mixer. Local drives on the qubits and the cavities steer and maintain the system to the desired qubit Bell state. Most spectacularly, even a weakly squeezed state can stabilize a maximally entangled Bell state of two distant qubits through an autonomous entanglement concentration process. Moreover, we show that such reservoir-engineering-based protocols can stabilize entanglement in the presence of qubit-cavity asymmetries and losses. This work was published in [12].

# 7. Partnerships and Cooperations

## 7.1. Regional Initiatives

- **Paris EMERGENCE project ENDURANCE**: In the framework of the Paris Ile de France program “EMERGENCE”, Zaki Leghtas has received a funding for his research program "Multi-photon processes in superconducting circuits for quantum error correction". This grant of 230k euros has allowed us to purchase the experimental equipment to complement the experiment based at ENS.
- **DIM SIRTEQ project Sputthy**: Zaki Leghtas has received 50k euros from the DIM SIRTEQ to purchase a sputtering system. With this machine, we will fabricate high quality resonators made out of Niobium and high kinetic inductance material such as NbTiN.
- **DIM SIRTEQ PhD fellowship**: We have received funding from DIM SIRTEQ to cover half of the PhD of Jérémie Guillaud under supervision of Mazyar Mirrahimi.
- **FSMP postdoctoral fellowship**: Paolo Forni has been selected for a postdoctoral fellowship by the Fondation des Sciences Mathématiques de Paris (FSMP) for the academic year 2018-2019: this 12-month postdoc fellowship extends a previous one supported by the programme Math-PSL of PSL Research University.

## 7.2. National Initiatives

- **ANR project GEARED**: This four-year collaborative ANR project, entitled “Reservoir engineering quantum entanglement in the microwave domain” and coordinated by Mazyar Mirrahimi, started on October 2014 and ended on September 2018. The participants of the project were Mazyar Mirrahimi (QUANTIC project-team), Benjamin Huard (ENS Lyon), Daniel Esteve and Fabien Portier (Quantronics group, CEA Saclay), Nicolas Roch and Olivier Buisson (Institut Neel, Grenoble). This project deals with robust generation of entanglement as a key resource for quantum information processing (quantum simulation, computation and communication). QUANTIC receieved a funding of 114k in this framework.
- **ANR project ENDURANCE**: In the framework of the ANR program “Accueil de chercheur de haut niveau”, Zaki Leghtas has received a funding for his research program "Multi-photon processes in superconducting circuits for quantum error correction". This grant of 400k euros has allowed us to purchase the experimental equipment to build a new experiment based at ENS. The project started in March 2016 for 42 months.
- **ANR project HAMROQS**: In the framework of the ANR program JCJC, Alain Sarlette has received a funding for his research program "High-accuracy model reduction for open quantum systems". This grant of 212k euros will start on april 2019 and will run for 4 years.

## 7.3. European Initiatives

### *7.3.1. Collaborations with Major European Organizations*

Partner 1: ENS Lyon

We are pursuing our interdisciplinary work about quantum control from theoretical aspects in direct collaboration with existing experiments (ENS Lyon) with the group of Benjamin Huard, former member of the QUANTIC team. Joint papers are published and underway. The ANR-JCJC project HAMROQS by Alain Sarlette has Benjamin Huard as external supporting collaborator.

Partner 2: Laboratoire Kastler Brossel

We have been collaborating with Samuel Deleglise and Emmanuel Flurin from Laboratoire Kastler Brossel to understand and analyze their experimental data. In this aim, we have developed new adiabatic elimination techniques for multi-partite open quantum systems with non-trivial zero-order dynamics.

Partner 3: University of Padova

Alain Sarlette has been pursuing a fruitful collaboration with the group of Francesco Ticozzi on "dynamical systems aspects of quantum systems". A novel line of work in the direction of quantum thermalization and quantum random walks has been explored, in the framework of the PhD of S. Apers (Ghent University) supervised by A. Sarlette.

Partner 4: Ghent University.

Alain Sarlette has been collaborating with applied mathematicians interested in quantum control at UGent (Dirk Aeyels, Lode Wylleman, Gert De Cooman) in the framework of thesis co-supervisions. Two PhD students have successfully defended their thesis this year (Arash Farnam, on distributed control of lattices; Simon Apers, on quantum walks). He is further coaching a Master thesis intern working on nonlinear deterministic structures in quantum SDEs.

## 7.4. International Initiatives

### *7.4.1. Inria Associate Teams Not Involved in an Inria International Labs*

TAQUILLA: is an Inria associate team (between Quantic team and Yale university) with principal Inria investigator, Mazyar Mirrahimi, and principal Yale investigator Michel Devoret. In this framework we continued our collaborations between Inria and Yale in 2018. Jérémie Guillaud visited Yale for 3 months (Sept-Nov), and Mazyar Mirrahimi for 4 months (Sept-Dec). Clarke Smith and Steven Touzard, PhD students at Yale, visited us for 1 week at the occasion of PRACQSYS meeting. Clarke Smith joins Quantic team as a postdoc in January 2019.

### *7.4.2. Participation in Other International Programs*

In the framework of the collaborations with Yale university, Quantic team has received a sub-award of 500k dollars over 4 years starting in 2018 from Yale university. This sub-award is part of an ARO (Army Research Office) grant received by our collaborators at Yale and covers the expenses related to our collaborations (hiring of new PhD students and postdocs at Inria and travels between Inria and Yale).

## 7.5. International Research Visitors

### *7.5.1. Visits of International Scientists*

- In the framework of Inria's invited professor program, Tryphon Georgiou (University of California at Irvine) visited us for about 2 months. This visit had for subject to initiate collaborations on the subject of open quantum systems and quantum channels.
- Yves Bérubé-Lauzière (University of Sherbrooke, Institut Quantique) accompanied by two PhD students made a 6-month visit from March to August 2018 to investigate with Pierre Rouchon feedback protocols for stabilizing quantum states in a high-quality cavity.
- P.S. Pereira da Silva (Escola Politécnica, PTC, University of SaoPaulo, Brazil) made a 2-week visit (June 25 to July 6) to investigate with Pierre Rouchon motion planning issues based on Lyapunov tracking for quantum gate generations.

### *7.5.2. Visits to International Teams*

*7.5.2.1. Research Stays Abroad*

In the framework of our collaborations with the group of Michel Devoret at Yale university, Jérémie Guillaud and Mazyar Mirrahimi visited Yale for 3 months and 4 months, respectively, in fall 2018.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. General Chair, Scientific Chair*

Pierre Rouchon was the main organizer of the spring thematic quarter at Institut Henri Poincaré entitled "Measurement and control of quantum systems: theory and experiments" (16 April – 13 July 2018). This thematic quarter includes courses, lectures and conferences. In particular a research school of one week at CIRM, two 3-day workshops in May and June and the 2018 issue of PRACQSYS conference in July, were organized in this framework. This thematic quarter involved several hundred of participants. See IHP web page (http://www.ihp.fr/en/CEB/T2-2018), CIRM web page (https://conferences.cirm-math.fr/1732.html) and the specific quarter web site (https://sites.google.com/view/mcqs2018/home).

*8.1.1.2. Member of the Organizing Committees*

- Zaki Leghtas co-organized a conference on quantum computing ICOQC 2018 at ENS Paris (https://icoqc.sciencesconf.org).
- Mazyar Mirrahimi was a co-organizer of the PRACQSYS conference at IHP in July 2018 (https://sites.google.com/view/mcqs2018/pracqsys-2018).

### *8.1.2. Journal*

*8.1.2.1. Member of the Editorial Boards*

Pierre Rouchon is member of the editorial board of Annual Reviews in Control.

*8.1.2.2. Reviewer - Reviewing Activities*

- Zaki Leghtas and Mazyar Mirrahimi were reviewer of Physical Review Journals.
- Pierre Rouchon and Alain Sarlette were reviewer for several automatic control and dynamical systems journals and conferences.

### *8.1.3. Invited Talks*

- Zaki Leghtas: Sirteq 2018. Institut d'optique Paris. Invited by Patrice Bertet.
- Zaki Leghtas: ONE-QOS workshop. Max Planck, Erlangen, Germany. Invited by Florian Marquardt.
- Zaki Leghtas: IHP workshop on quantum control and feedback, Paris, France. Invited by Eleni Diamanti.
- Zaki Leghtas: LIA CNRS-Université de Sherbrooke workshop, Saint-Rémy, France. Invited by Denis Vion.
- Zaki Leghtas: University of Berkeley. Berkeley, USA. Invited by Irfan Siddiqi.
- Zaki Leghtas: Rigetti Quantum Computing. Berkeley, USA. Invited by Chad Rigetti.
- Pierre Rouchon: lecture at the QUACO ANR Meeting in Besançon, September 24-26, Models and feedback issues for open quantum systems.

- Pierre Rouchon: plenary speaker at Mexican Annual Conference on Automatic Control. 10-12 October 2018, San Luis Potosi, Dynamical models and feedback issues for super-conducting quantum circuits.
- Pierre Rouchon: 2-hour course in the Colloquium of the Physics Department, ENS-Paris, October 23 (introduction to quantum cryptography, computation and error correction).
- Alain Sarlette: dynamical systems seminar series, March 2018, Jussieu.
- Alain Sarlette: seminar at IHP trimester on Quantum Control, May 2018.
- Alain Sarlette: Praqcsys: Principles and Applications of Control in Quantum Systems, IHP, Paris, July 2018.
- Mazyar Mirrahimi: American Physical Society March Meeting, Los Angeles, March 2018.
- Mazyar Mirrahimi: Semi-plenary speaker at MTNS (Mathematical Theory of Networks and Systems), Hong Kong, July 2018.
- Mazyar Mirrahimi: Centre de Recherche Mathematique de Montreal, Octobre 2018.
- Mazyar Mirrahimi: 4-hour course at Institut d'Optique, Introduction to Quantum Computing, June 2018.
- Alain Sarlette: lectures on quantum control and quantum computing at the Ecole d'Automatique de Grenoble summer school, August 2018
- Pierre Rouchon, Alain Sarlette, Rémi Azouit, Paolo Forni and Francesca Chittaro have given a lecture series about "adiabatic elimination for open quantum systems" at the IHP trimester on Quantum Control.
- Jérémie Guillaud: Yale university, Nov 2018.

### *8.1.4. Research Administration*

- Pierre Rouchon is a member of the scientific committee of LAGEP (Laboratoire d'Automatique et de Génie des Procédés) since 2017.
- Pierre Rouchon is a membre of the "Conseil Scientifique du DIM Math Innov" since 2017.
- Pierre Rouchon is a member of the "Conseil de la recherche de PSL " since 2016.
- Pierre Rouchon is a member of the "Conseil Scientifique du Conservatoire National des Arts et Metiers" since 2014.
- Mazyar Mirrahimi is the co-president of Inria's comité des emplois scientifiques.
- Mazyar Mirrahimi was a member of ANR Comité d'Evaluation Scientifique on Quantum Technologies.

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Cycle Ingénieur : Mazyar Mirrahimi, Automatic Control with Applications in Robotics and in Quantum Engineering, 8 hours amphi and 8 hours TD, 3rd year, Ecole Polytechnique, France.

Cycle Ingénieur : Mazyar Mirrahimi, Contrôle de modèles dynamiques, 36 hours TD, 2nd year, Ecole Polytechnique, France.

Cycle Ingénieur : Mazyar Mirrahimi, Module algorithmique Quantum Control, 24 hours TD, 2nd year, Ecole Polytechnique, France.

Master: Mazyar Mirrahimi and Pierre Rouchon, Dynamics and control of quantum systems, 18 hours amphi, M2, Jussieu, France.

Cycle Ingénieur: Alain Sarlette, Probabilities and Stochastic Processes, 24 hours TD, Mines Paristech, France.

Master: Alain Sarlette, Robotics, 24 hours, Ghent University, Belgium.

Cycle Ingénieur : Zaki Leghtas, Quantum Mechanics and Statistical Physics, Mines ParisTech, 12 hours, France.

Cycle Ingénieur : Zaki Leghtas, Quantum Computing, Mines ParisTech, 20 hours, France.

### *8.2.2. Supervision*

- PhD in progress : Gerardo Cardona, "Beyond static gains in analog quantum feedback control", advisors: Pierre Rouchon and Alain Sarlette, starting date: Nov 2016.
- PhD in progress: Michiel Burgelman, "A systematic study of strongly driven and dissipative quantum systems towards high-accuracy quantum control designs", advisors: Pierre Rouchon and Alain Sarlette, starting date: Nov 2018.
- PhD in progress: Vincent Martin, "Fault-tolerance of quantum systems under continuous-time feedback stabilization", advisors: Mazyar Mirrahimi and Alain Sarlette, starting date Oct 2018.
- PhD in progress: Jérémie Guillaud, "Modular architecture for quantum information processing", advisor: Mazyar Mirrahimi and Pierre Rouchon, starting date Oct 2017.
- PhD in progress: Lucas Verney, "Robust processing of quantum information with superconducting circuits", advisor: Mazyar Mirrahimi and Zaki Leghtas, starting date Oct 2016.
- PhD in progress: Raphaël Lescanne, “Engineering Multi-Photon Dissipation In Superconducting Circuits For Quantum Error Correction”, advisors: Zaki Leghtas and Takis Kontos, starting date Sept 2016.
- PhD in progress: Marius Villiers, “Probing the spin entanglement of single Cooper pair”, advisors: Zaki Leghtas and Takis Kontos, starting date: September 2018.
- Alain Sarlette has been supervising 2 PhD students with his former institution UGent. Arash Farnam has successfully defended his thesis about distributed systems control in October 2018. Simon Apers has successfully defended his thesis about quantum walks on graphs in November 2018.

### *8.2.3. Juries*

Pierre Rouchon was the president of the jury for the Habilitation thesis of Fransceca Chirraro (université de Toulon) and member of the jury for the Habilitation thesis of Nadir Farhi (université Paris-Est).

## 8.3. Popularization

Alain Sarlette, 5 December 2018, prospective Ordinateur Quantique at the comité de pilotage du CETIM, Senlis.

# 9. Bibliography

## Major publications by the team in recent years

[1] H. AMINI, A. SOMARAJU, I. DOTSENKO, C. SAYRIN, M. MIRRAHIMI, P. ROUCHON.*Feedback stabilization of discrete-time quantum systems subject to non-demolition measurements with imperfections and delays*, in "Automatica", 2013, vol. 49, n$^{o}$ 9, p. 2683–2692

[2] P. CAMPAGNE-IBARCQ, P. SIX, L. BRETHEAU, A. SARLETTE, M. MIRRAHIMI, P. ROUCHON, B. HUARD.*Observing Quantum State Diffusion by Heterodyne Detection of Fluorescence* , in "Physical Review X", January 2016, vol. 6, 011002 [*DOI :* 10.1103/PHYSREVX.6.011002], https://hal-mines-paristech.archives-ouvertes.fr/hal-01264326

[3] J. COHEN, W. C. SMITH, M. H. DEVORET, M. MIRRAHIMI.*Degeneracy-preserving quantum non-demolition measurement of parity-type observables for cat-qubits*, in "Physical Review Letters", August 2017, 25 pages, 7 figures [*DOI :* 10.1103/PHYSREVLETT.119.060503], https://hal.inria.fr/hal-01437156

[4] Z. LEGHTAS, S. TOUZARD, I. M. POP, A. KOU, B. VLASTAKIS, A. PETRENKO, K. M. SLIWA, A. NARLA, S. SHANKAR, M. J. HATRIDGE, M. REAGOR, L. FRUNZIO, R. J. SCHOELKOPF, M. MIRRAHIMI, M. H. DEVORET.*Confining the state of light to a quantum manifold by engineered two-photon loss*, in "Science", February 2015, vol. 347, n$^{o}$ 6224, p. 853-857 [*DOI :* 10.1126/SCIENCE.AAA2085], https://hal.inria.fr/hal-01240210

[5] N. OFEK, A. PETRENKO, R. HEERES, P. REINHOLD, Z. LEGHTAS, B. VLASTAKIS, Y. LIU, L. FRUNZIO, S. GIRVIN, L. JIANG, M. MIRRAHIMI, M. H. DEVORET, R. J. SCHOELKOPF.*Extending the lifetime of a quantum bit with error correction in superconducting circuits*, in "Nature", 2016, vol. 536, 5

[6] S. ROSENBLUM, P. REINHOLD, M. MIRRAHIMI, L. JIANG, L. FRUNZIO, R. J. SCHOELKOPF.*Fault-tolerant detection of a quantum error*, in "Science", July 2018, vol. 361, n$^{o}$ 6399, p. 266 - 270, https://arxiv.org/abs/1803.00102 [*DOI :* 10.1126/SCIENCE.AAT3996], https://hal.inria.fr/hal-01929080

[7] A. SARLETTE, J.-M. RAIMOND, M. BRUNE, P. ROUCHON.*Stabilization of nonclassical states of the radiation field in a cavity by reservoir engineering*, in "Phys. Rev. Lett.", 2011, vol. 107, 010402

[8] C. SAYRIN, I. DOTSENKO, X. ZHOU, B. PEAUDECERF, T. RYBARCZYK, S. GLEYZES, P. ROUCHON, M. MIRRAHIMI, H. AMINI, M. BRUNE, J.-M. RAIMOND, S. HAROCHE.*Real-time quantum feedback prepares and stabilizes photon number states*, in "Nature", 2011, vol. 477, p. 73–77

[9] S. SHANKAR, M. HATRIDGE, Z. LEGHTAS, K. SLIWA, A. NARLA, U. VOOL, S. GIRVIN, L. FRUNZIO, M. MIRRAHIMI, M. H. DEVORET.*Autonomously stabilized entanglement between two superconducting quantum bits*, in "Nature", 2013, vol. 504, p. 419–422

[10] C. WANG, Y. GAO, P. REINHOLD, R. HEERES, N. OFEK, K. CHOU, C. AXLINE, M. REAGOR, J. BLUMOFF, K. SLIWA, L. FRUNZIO, S. GIRVIN, L. JIANG, M. MIRRAHIMI, M. H. DEVORET, R. J. SCHOELKOPF.*A Schrodinger cat living in two boxes*, in "Science", 2016, vol. 352, 5

## Publications of the year

### Articles in International Peer-Reviewed Journal

[11] S. APERS, A. SARLETTE, F. TICOZZI.*Simulation of quantum walks and fast mixing with classical processes*, in "Physical Review A", September 2018, vol. 98, n$^{o}$ 3, 032115, https://hal.inria.fr/hal-01946763

[12] N. DIDIER, J. GUILLAUD, S. SHANKAR, M. MIRRAHIMI.*Remote entanglement stabilization and concentration by quantum reservoir engineering*, in "Physical Review A", July 2018, https://arxiv.org/abs/1703.03379 - 5 pages, 4 figures [*DOI :* 10.1103/PHYSREVA.98.012329], https://hal.inria.fr/hal-01652766

[13] Q. FICHEUX, S. JEZOUIN, Z. LEGHTAS, B. HUARD.*Dynamics of a qubit while simultaneously monitoring its relaxation and dephasing*, in "Nature Communications", May 2018, vol. 9, n$^{o}$ 1926, https://arxiv.org/abs/1711.01208 - Supplemental videos can be found at http://physinfo.fr/publications/Ficheux1710.html and supplemental information can be found as an ancillary file on arxiv [*DOI :* 10.1038/S41467-018-04372-9], https://hal.archives-ouvertes.fr/hal-01779089

[14] D. MARKOVIĆ, S. JEZOUIN, Q. FICHEUX, S. FEDORTCHENKO, S. FELICETTI, T. COUDREAU, P. MILMAN, Z. LEGHTAS, B. HUARD.*Demonstration of an Effective Ultrastrong Coupling between Two Oscillators*, in "Physical Review Letters", July 2018, vol. 121, n$^{o}$ 4 [*DOI :* 10.1103/PHYSREVLETT.121.040505], https://hal-mines-paristech.archives-ouvertes.fr/hal-01944732

[15] P. MARTIN, L. ROSIER, P. ROUCHON.*Controllability of the 1D Schrödinger equation using flatness*, in "Automatica", May 2018, vol. 91, p. 208 - 216 [*DOI :* 10.1016/J.AUTOMATICA.2018.01.005], https://hal-mines-paristech.archives-ouvertes.fr/hal-01769227

[16] S. ROSENBLUM, Y. GAO, P. REINHOLD, C. WANG, C. AXLINE, L. FRUNZIO, S. GIRVIN, L. JIANG, M. MIRRAHIMI, M. H. DEVORET, R. J. SCHOELKOPF.*A CNOT gate between multiphoton qubits encoded in two cavities*, in "Nature Communications", February 2018, https://arxiv.org/abs/1709.05425 - 10 pages, 11 figures (incl. Supplementary Information) [*DOI :* 10.1038/S41467-018-03059-5], https://hal.inria.fr/hal-01652773

[17] S. ROSENBLUM, P. REINHOLD, M. MIRRAHIMI, L. JIANG, L. FRUNZIO, R. J. SCHOELKOPF.*Fault-tolerant detection of a quantum error*, in "Science", July 2018, vol. 361, n$^{o}$ 6399, p. 266 - 270, https://arxiv.org/abs/1803.00102 [*DOI :* 10.1126/SCIENCE.AAT3996], https://hal.inria.fr/hal-01929080

[18] S. TOUZARD, A. GRIMM, Z. LEGHTAS, S. O. MUNDHADA, P. REINHOLD, C. AXLINE, M. REAGOR, K. CHOU, J. BLUMOFF, K. M. SLIWA, S. SHANKAR, L. FRUNZIO, R. J. SCHOELKOPF, M. MIRRAHIMI, M. H. DEVORET.*Coherent oscillations inside a quantum manifold stabilized by dissipation*, in "Physical Review X", April 2018, https://arxiv.org/abs/1705.02401 [*DOI :* 10.1103/PHYSREVX.8.021005], https://hal.inria.fr/hal-01652771

### International Conferences with Proceedings

[19] G. CARDONA, A. SARLETTE, P. ROUCHON.*Exponential stochastic stabilization of a two-level quantum system via strict Lyapunov control*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Miami, United States, December 2018, https://hal.inria.fr/hal-01946780

[20] P. FORNI, A. SARLETTE, T. CAPELLE, E. FLURIN, S. DELÉGLISE, P. ROUCHON.*Adiabatic elimination for multi-partite open quantum systems with non-trivial zero-order dynamics*, in "CDC 2018 - 57th IEEE Conference on Decision and Control", Miami, United States, December 2018, https://hal.inria.fr/hal-01946773

### Other Publications

[21] R. LESCANNE, L. VERNEY, Q. FICHEUX, M. H. DEVORET, B. HUARD, M. MIRRAHIMI, Z. LEGHTAS.*Dynamics of an off-resonantly pumped superconducting qubit in a cavity*, December 2018, https://arxiv.org/abs/1805.05198 - working paper or preprint, https://hal.inria.fr/hal-01945882

[22] Z. K. MINEV, S. O. MUNDHADA, S. SHANKAR, P. REINHOLD, R. GUTIERREZ-JAUREGUI, R. J. SCHOELKOPF, M. MIRRAHIMI, H. J. CARMICHAEL, M. H. DEVORET.*To catch and reverse a quantum jump mid-flight*, November 2018, https://arxiv.org/abs/1803.00545 - Added sections to the Supplementary Information on the concepts and signal-to-noise ratio of the experiment. Corrected a few minor typos and clarified text. Revised and expanded citations, https://hal.inria.fr/hal-01929083

[23] S. O. MUNDHADA, A. GRIMM, J. VENKATRAMAN, Z. K. MINEV, S. TOUZARD, N. E. FRATTINI, V. V. SIVAK, K. SLIWA, P. REINHOLD, S. SHANKAR, M. MIRRAHIMI, M. H. DEVORET.*Experimental implementation of a Raman-assisted six-quanta process*, November 2018, https://arxiv.org/abs/1811.06589 - 13 pages, 5 figures, https://hal.inria.fr/hal-01936696

[24] S. PURI, A. GRIMM, P. CAMPAGNE-IBARCQ, A. EICKBUSCH, K. NOH, G. ROBERTS, L. JIANG, M. MIRRAHIMI, M. H. DEVORET, S. M. GIRVIN.*Stabilized Cat in Driven Nonlinear Cavity: A Fault-Tolerant Error Syndrome Detector*, November 2018, https://arxiv.org/abs/1807.09334 - working paper or preprint, https://hal.inria.fr/hal-01936700

[25] L. VERNEY, R. LESCANNE, M. H. DEVORET, Z. LEGHTAS, M. MIRRAHIMI.*Strongly driven quantum Josephson circuits*, December 2018, https://arxiv.org/abs/1805.07542 - 14 pages, 9 figures, https://hal.inria.fr/hal-01945870

## References in notes

[26] S. ATTAL, A. JOYE, C.-A. PILLET (editors). *Open Quantum Systems III: Recent Developments*, Springer, Lecture notes in Mathematics 1880, 2006

[27] H. AMINI, M. MIRRAHIMI, P. ROUCHON.*Stabilization of a delayed quantum system: the Photon Box case-study*, in "IEEE Trans. Automatic Control", 2012, vol. 57, n$^{o}$ 8, p. 1918–1930

[28] H. AMINI, C. PELLEGRINI, P. ROUCHON.*Stability of continuous-time quantum filters with measurement imperfections*, in "Russian Journal of Mathematical Physics", 2014, vol. 21, p. 297–315

[29] H. AMINI, A. SOMARAJU, I. DOTSENKO, C. SAYRIN, M. MIRRAHIMI, P. ROUCHON.*Feedback stabilization of discrete-time quantum systems subject to non-demolition measurements with imperfections and delays*, in "Automatica", 2013, vol. 49, n$^{o}$ 9, p. 2683–2692

[30] A. BARCHIELLI, M. GREGORATTI.*Quantum Trajectories and Measurements in Continuous Time: the Diffusive Case*, Springer Verlag, 2009

[31] J. BARREIRO, M. MULLER, P. SCHINDLER, D. NIGG, T. MONZ, M. CHWALLA, M. HENNRICH, C. ROOS, P. ZOLLER, R. BLATT.*An open-system quantum simulator with trapped ions*, in "Nature", 2011, vol. 470, 486

[32] V. BELAVKIN.*Quantum stochastic calculus and quantum nonlinear filtering*, in "Journal of Multivariate Analysis", 1992, vol. 42, n$^{o}$ 2, p. 171–201

[33] T. BENOIST, C. PELLEGRINI.*Large Time Behavior and Convergence Rate for Quantum Filters Under Standard Non Demolition Conditions*, in "Communications in Mathematical Physics", 2014, p. 1-21, http://dx.doi.org/10.1007/s00220-014-2029-6

[34] G. BIRKHOFF.*Extensions of Jentzch's theorem*, in "Trans. Amer. Math. Soc.", 1957, vol. 85, p. 219–227

[35] S. BOLOGNANI, F. TICOZZI.*Engineering stable discrete-time quantum dynamics via a canonical QR decomposition*, in "IEEE Trans. Autom. Control", 2010, vol. 55

[36] V. BRAGINSKI, F. KHALILI.*Quantum Measurements*, Cambridge University Press, 1992

[37] P. CAMPAGNE-IBARCQ, L. BRETHEAU, E. FLURIN, A. AUFFÈVES, F. MALLET, B. HUARD.*Observing Interferences between Past and Future Quantum States in Resonance Fluorescence*, in "Phys. Rev. Lett.", May 2014, vol. 112, 180402, http://link.aps.org/doi/10.1103/PhysRevLett.112.180402

[38] P. CAMPAGNE-IBARCQ, E. FLURIN, N. ROCH, D. DARSON, P. MORFIN, M. MIRRAHIMI, M. H. DEVORET, F. MALLET, B. HUARD.*Persistent Control of a Superconducting Qubit by Stroboscopic Measurement Feedback*, in "Phys. Rev. X", 2013, vol. 3, 021008

[39] H. CARMICHAEL.*Statistical Methods in Quantum Optics 2: Non-Classical Fields*, Spinger, 2007

[40] H. CARMICHAEL.*An Open Systems Approach to Quantum Optics*, Springer-Verlag, 1993

[41] J. CARR.*Application of Center Manifold Theory*, Springer, 1981

[42] J. COHEN, M. MIRRAHIMI.*Dissipation-induced continuous quantum error correction for superconducting circuits*, in "Phys. Rev. A", 2014, vol. 90, 062344

[43] J. DALIBARD, Y. CASTIN, K. MÖLMER.*Wave-function approach to dissipative processes in quantum optics*, in "Phys. Rev. Lett.", 1992, vol. 68, n$^{o}$ 5, p. 580–583

[44] M. H. DEVORET, A. WALLRAFF, J. MARTINIS.*Superconducting Qubits: A Short Review*, 2004, arXiv:cond-mat/0411174

[45] I. DOTSENKO, M. MIRRAHIMI, M. BRUNE, S. HAROCHE, J.-M. RAIMOND, P. ROUCHON.*Quantum feedback by discrete quantum non-demolition measurements: towards on-demand generation of photon-number states*, in "Physical Review A", 2009, vol. 80: 013805-013813

[46] N. FENICHEL.*Geometric singular perturbation theory for ordinary differential equations*, in "J. Diff. Equations", 1979, vol. 31, p. 53–98

[47] J. GAMBETTA, H. M. WISEMAN.*State and dynamical parameter estimation for open quantum systems*, in "Phys. Rev. A", September 2001, vol. 64, n$^{o}$ 4, 042105, http://link.aps.org/doi/10.1103/PhysRevA.64.042105

[48] S. GAMMELMARK, B. JULSGAARD, K. MÖLMER.*Past Quantum States of a Monitored System*, in "Phys. Rev. Lett.", October 2013, vol. 111, n$^{o}$ 16, 160401, http://link.aps.org/doi/10.1103/PhysRevLett.111.160401

[49] C. GARDINER, P. ZOLLER.*Quantum Noise*, third, Springer, 2010

[50] S. GAUBERT, Z. QU.*Checking the strict positivity of Kraus maps is NP-hard*, in "arXiv:1402.1429", 2014

[51] S. GAUBERT, Z. QU.*The contraction rate in Thompson's part metric of order-preserving flows on a cone - Application to generalized Riccati equations*, in "Journal of Differential Equations", April 2014, vol. 256, n$^{o}$ 8, p. 2902–2948, http://www.sciencedirect.com/science/article/pii/S0022039614000424

[52] K. GEERLINGS, Z. LEGHTAS, I. POP, S. SHANKAR, L. FRUNZIO, R. SCHOELKOPF, M. MIRRAHIMI, M. H. DEVORET.*Demonstrating a Driven Reset Protocol of a Superconducting Qubit*, in "Phys. Rev. Lett.", 2013, vol. 110, 120501

[53] D. GOTTESMAN, A. KITAEV, J. PRESKILL.*Encoding a qubit in an oscillator*, in "Phys. Rev. A", 2001, vol. 64, 012310

[54] C. GUERLIN, J. BERNU, S. DELÉGLISE, C. SAYRIN, S. GLEYZES, S. KUHR, M. BRUNE, J.-M. RAIMOND, S. HAROCHE.*Progressive field-state collapse and quantum non-demolition photon counting*, in "Nature", 2007, vol. 448, p. 889-893

[55] S. HAROCHE, J.-M. RAIMOND.*Exploring the Quantum: Atoms, Cavities and Photons*, Oxford University Press, 2006

[56] M. HATRIDGE, S. SHANKAR, M. MIRRAHIMI, F. SCHACKERT, K. GEERLINGS, T. BRECHT, K. SLIWA, B. ABDO, L. FRUNZIO, S. GIRVIN, R. SCHOELKOPF, M. H. DEVORET.*Quantum back-action of an individual variable-strength measurement*, in "Science", 2013, vol. 339, p. 178–181

[57] E. HOLLAND, B. VLASTAKIS, R. HEERES, M. REAGOR, U. VOOL, Z. LEGHTAS, L. FRUNZIO, G. KIRCHMAIR, M. DEVORET, M. MIRRAHIMI, R. SCHOELKOPF.*Single-photon-resolved cross-Kerr interaction for autonomous stabilization of photon-number states*, in "Phys. Rev. Lett.", 2015, vol. 115, 180501

[58] T. KATO.*Perturbation Theory for Linear Operators*, Springer, 1966

[59] E. KNILL, R. LAFLAMME, G. MILBURN.*A scheme for efficient quantum computation with linear optics*, in "Nature", 2001, vol. 409, 46

[60] H. KRAUTER, C. MUSCHIK, K. JENSEN, W. WASILEWSKI, J. PETERSEN, J. CIRAC, E. POLZIK.*Entanglement Generated by Dissipation and Steady State Entanglement of Two Macroscopic Objects*, in "Phys. Rev. Lett.", 2011, vol. 107, 080503

[61] Z. LEGHTAS, G. KIRCHMAIR, B. VLASTAKIS, M. H. DEVORET, R. J. SCHOELKOPF, M. MIRRAHIMI.*Deterministic protocol for mapping a qubit to coherent state superpositions in a cavity*, in "Phys. Rev. A", 2013, vol. 87, 042315

[62] Z. LEGHTAS, G. KIRCHMAIR, B. VLASTAKIS, R. J. SCHOELKOPF, M. H. DEVORET, M. MIRRAHIMI.*Hardware-efficient autonomous quantum memory protection*, in "Phys. Rev. Lett.", 2013, vol. 111, 120501

[63] Z. LEGHTAS, A. SARLETTE, P. ROUCHON.*Adiabatic passage and ensemble control of quantum systems*, in "J. Phys. B", 2011, vol. 44, 154017

[64] Z. LEGHTAS, U. VOOL, S. SHANKAR, M. HATRIDGE, S. GIRVIN, M. H. DEVORET, M. MIRRAHIMI.*Stabilizing a Bell state of two superconducting qubits by dissipation engineering*, in "Phys. Rev. A", 2013, vol. 88, 023849

[65] J.-S. LI, N. KHANEJA.*Ensemble control of Bloch equations*, in "IEEE Trans. Autom. Control", 2009, vol. 54, p. 528–536

[66] Y. LIN, J. GAEBLER, F. REITER, T. TAN, R. BOWLER, A. SORENSEN, D. LEIBFRIED, D. WINELAND.*Dissipative production of a maximally entangled steady state of two quantum bits*, in "Nature", 2013, vol. 504, p. 415–418

[67] S. LLOYD.*Coherent quantum feedback*, in "Phys. Rev. A", 2000, vol. 62, 022108

[68] L. MAZZARELLA, A. SARLETTE, F. TICOZZI.*Consensus for quantum networks: from symmetry to gossip iterations*, in "IEEE Trans. Automat. Control", 2014, in press

[69] M. MIRRAHIMI, B. HUARD, M. H. DEVORET.*Strong measurement and quantum feedback for persistent Rabi oscillations in circuit QED experiments*, in "IEEE Conference on Decision and Control", IEEE Conference on Decision and Control, 2012

[70] M. MIRRAHIMI, Z. LEGHTAS, V. ALBERT, S. TOUZARD, R. J. SCHOELKOPF, L. JIANG, M. H. DEVORET.*Dynamically protected cat-qubits: a new paradigm for universal quantum computation*, in "New J. Phys.", 2014, vol. 16, 045014

[71] K. MURCH, U. VOOL, D. ZHOU, S. WEBER, S. GIRVIN, I. SIDDIQI.*Cavity-assisted quantum bath engineering*, in "Phys. Rev. Lett.", 2012, vol. 109, 183602

[72] A. NEGRETTI, K. MÖLMER.*Estimation of classical parameters via continuous probing of complementary quantum observables*, in "New Journal of Physics", 2013, vol. 15, n$^{o}$ 12, 125002, http://stacks.iop.org/1367-2630/15/i=12/a=125002

[73] H. NURDIN, M. JAMES, I. PETERSEN.*Coherent quantum LQG control*, in "Automatica", 2009, vol. 45, p. 1837–1846

[74] B. PEAUDECERF, T. RYBARCZYK, S. GERLICH, S. GLEYZES, J.-M. RAIMOND, S. HAROCHE, I. DOTSENKO, M. BRUNE.*Adaptive Quantum Nondemolition Measurement of a Photon Number*, in "Phys. Rev. Lett.", Feb 2014, vol. 112, n$^{o}$ 8, 080401, http://link.aps.org/doi/10.1103/PhysRevLett.112.080401

[75] D. PETZ.*Monotone Metrics on matrix spaces*, in "Linear Algebra and its Applications", 1996, vol. 244, p. 81–96

[76] J. POYATOS, J. CIRAC, P. ZOLLER.*Quantum Reservoir Engineering with Laser Cooled Trapped Ions*, in "Phys. Rev. Lett.", 1996, vol. 77, n$^{o}$ 23, p. 4728–4731

[77] D. REEB, M. J. KASTORYANO, M. M. WOLF.*Hilbert's projective metric in quantum information theory*, in "Journal of Mathematical Physics", August 2011, vol. 52, n$^{o}$ 8, 082201, http://dx.doi.org/10.1063/1.3615729

[78] D. RISTÈ, J. LEEUWEN, H.-S. KU, K. LEHNERT, L. DICARLO.*Initialization by measurement of a superconducting quantum bit circuit*, in "Phys. Rev. Lett.", 2012, vol. 109, 050507

[79] N. ROCH, E. FLURIN, F. NGUYEN, P. MORFIN, P. CAMPAGNE-IBARCQ, M. H. DEVORET, B. HUARD.*Widely tunable, non-degenerate three-wave mixing microwave device operating near the quantum limit*, in "Phys. Rev. Lett.", 2012, vol. 108, 147701

[80] P. ROUCHON.*Fidelity is a Sub-Martingale for Discrete-Time Quantum Filters*, in "IEEE Transactions on Automatic Control", 2011, vol. 56, n$^{o}$ 11, p. 2743–2747

[81] A. ROY, Z. LEGHTAS, A. STONE, M. DEVORET, M. MIRRAHIMI.*Continuous generation and stabilization of mesoscopic field superposition states in a quantum circuit*, in "Phys. Rev. A", 2015, vol. 91, 013810

[82] A. SARLETTE, Z. LEGHTAS, M. BRUNE, J.-M. RAIMOND, P. ROUCHON.*Stabilization of nonclassical states of one- and two-mode radiation fields by reservoir engineering*, in "Phys. Rev. A", 2012, vol. 86, 012114

[83] D. SCHUSTER, A. HOUCK, J. SCHREIER, A. WALLRAFF, J. GAMBETTA, A. BLAIS, L. FRUNZIO, J. MAJER, B. JOHNSON, M. H. DEVORET, S. GIRVIN, R. J. SCHOELKOPF.*Resolving photon number states in a superconducting circuit*, in "Nature", 2007, vol. 445, p. 515–518

[84] R. SEPULCHRE, A. SARLETTE, P. ROUCHON.*Consensus in non-commutative spaces*, in "Decision and Control (CDC), 2010 49th IEEE Conference on", 2010, p. 6596–6601

[85] P. SHOR.*Scheme for reducing decoherence in quantum memory*, in "Phys. Rev. A", 1995, vol. 52, p. 2493–2496

[86] A. SOMARAJU, I. DOTSENKO, C. SAYRIN, P. ROUCHON.*Design and Stability of Discrete-Time Quantum Filters with Measurement Imperfections*, in "American Control Conference", 2012, p. 5084–5089

[87] A. SOMARAJU, M. MIRRAHIMI, P. ROUCHON.*Approximate stabilization of infinite dimensional quantum stochastic system*, in "Reviews in Mathematical Physics", 2013, vol. 25, 1350001

[88] A. STEANE.*Error Correcting Codes in Quantum Theory*, in "Phys. Rev. Lett", 1996, vol. 77, n$^{o}$ 5

[89] L. SUN, A. PETRENKO, Z. LEGHTAS, B. VLASTAKIS, G. KIRCHMAIR, K. SLIWA, A. NARLA, M. HATRIDGE, S. SHANKAR, J. BLUMOFF, L. FRUNZIO, M. MIRRAHIMI, M. H. DEVORET, R. J. SCHOELKOPF.*Tracking photon jumps with repeated quantum non-demolition parity measurements*, in "Nature", 2014, vol. 511, p. 444–448

[90] J. TSITSIKLIS.*Problems in decentralized decision making and computation*, in "PhD Thesis, MIT", 1984

[91] R. VIJAY, C. MACKLIN, D. SLICHTER, S. WEBER, K. MURCH, R. NAIK, A. KOROTKOV, I. SIDDIQI.*Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback*, in "Nature", 2012, vol. 490, p. 77–80

[92] L. VIOLA, E. KNILL, S. LLOYD.*Dynamical decoupling of open quantum system*, in "Phys. Rev. Lett.", 1999, vol. 82, p. 2417-2421

[93] B. VLASTAKIS, G. KIRCHMAIR, Z. LEGHTAS, S. NIGG, L. FRUNZIO, S. GIRVIN, M. MIRRAHIMI, M. H. DEVORET, R. J. SCHOELKOPF.*Deterministically encoding quantum information using 100-photon Schrödinger cat states*, in "Science", 2013, vol. 342, p. 607–610

[94] X. ZHOU, I. DOTSENKO, B. PEAUDECERF, T. RYBARCZYK, C. SAYRIN, S. GLEYZES, J.-M. RAIMOND, M. BRUNE, S. HAROCHE.*Field locked to Fock state by quantum feedback with single photon corrections*, in "Physical Review Letter", 2012, vol. 108, 243602

[95] R. VAN HANDEL.*The stability of quantum Markov filters*, in "Infin. Dimens. Anal. Quantum Probab. Relat. Top.", 2009, vol. 12, p. 153–172

# Project-Team REO

## Numerical simulation of biological flows

IN COLLABORATION WITH: Laboratoire Jacques-Louis Lions (LJLL)

IN PARTNERSHIP WITH:
**CNRS**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Modeling and Control for Life Sciences**

# Table of contents

# Project-Team REO

*Creation of the Project-Team: 2005 April 01, end of the Project-Team: 2018 December 31*

**Keywords:**

**Computer Science and Digital Science:**

A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.3.1. - Inverse problems
A6.3.2. - Data assimilation
A6.3.4. - Model reduction

**Other Research Topics and Application Domains:**

B2.2.1. - Cardiovascular and respiratory diseases
B2.2.3. - Cancer
B2.4.1. - Pharmaco kinetics and dynamics

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Miguel Ángel Fernández Varela [Team leader, Inria, Senior Researcher, HDR]
Jean-Frédéric Gerbeau [Team leader, Inria, Senior Researcher, until Jul 2018, HDR]
Céline Grandmont [Inria, Senior Researcher, HDR]
Damiano Lombardi [Inria, Researcher]
Marc Thiriet [CNRS, Researcher, HDR]
Marina Vidrascu [Inria, Emeritus]
Irene Vignon Clementel [Inria, Senior Researcher]

**Faculty Members**

Laurent Boudin [Univ Pierre et Marie Curie, Associate Professor]
Muriel Boulakia [Univ Pierre et Marie Curie, Associate Professor]

**Technical Staff**

Gautier Bureau [Inria, until Feb 2018]
Daniele Carlo Corti [Inria, from Oct 2018]
Fabien Raphel [Inria]

**PhD Students**

Ludovic Boilevin-Kayl [Inria]
Chen-Yu Chiang [Univ Pierre et Marie Curie]
Felipe Galarce Marin [Inria]
Fannie Gerosa [Inria]
Nicolas Golse [Inria-APHP, from Nov 2018]
Alexandre This [Philips, PhD Student, granted by CIFRE]

**Post-Doctoral Fellows**

Jean Jerome Casanova [Inria, from Sep 2018]
Florian Joly [Inria, until Jul 2018]

**Visiting Scientist**

Marthe Combari [Univ. Paris-Saclay, until Oct 2018]

**Administrative Assistant**

Maryse Desnous [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

REO is a joint project-team of the Inria Research Center of Paris and the Jacques-Louis Lions Laboratory (LJLL) of the Pierre and Marie Curie University (Sorbonne Université, UPMC Paris 6) and CNRS (UMR7598). Its main objectives are:

- the modeling of blood flow in large vessels, air flow in the respiratory tract, and the cardiac electrophysiology;
- the design and the analysis of efficient and robust numerical methods for these problems;
- the development of numerical software to assist medical decisions and to contribute to the design of medical devices.

REO put a strong effort in working with real data, coming either from clinicians or industrial partners. The development of methods for the interaction of data and simulation is therefore an important aspect of the activity of the team.

# 3. Research Program

## 3.1. Multiphysics modeling

In large vessels and in large bronchi, blood and air flows are generally supposed to be governed by the incompressible Navier-Stokes equations. Indeed in large arteries, blood can be supposed to be Newtonian, and at rest air can be modeled as an incompressible fluid. The cornerstone of the simulations is therefore a Navier-Stokes solver. But other physical features have also to be taken into account in simulations of biological flows, in particular fluid-structure interaction in large vessels and transport of sprays, particles or chemical species.

### *3.1.1. Fluid-structure interaction*

Fluid-structure coupling occurs both in the respiratory and in the circulatory systems. We focus mainly on blood flows since our work is more advanced in this field. But the methods developed for blood flows could be also applied to the respiratory system.

Here “fluid-structure interaction” means a coupling between the 3D Navier-Stokes equations and a 3D (possibly thin) structure in large displacements.

The numerical simulations of the interaction between the artery wall and the blood flows raise many issues: (1) the displacement of the wall cannot be supposed to be infinitesimal, geometrical nonlinearities are therefore present in the structure and the fluid problem have to be solved on a moving domain (2) the densities of the artery walls and the blood being close, the coupling is strong and has to be tackled very carefully to avoid numerical instabilities, (3) “naive” boundary conditions on the artificial boundaries induce spurious reflection phenomena.

Simulation of valves, either at the outflow of the cardiac chambers or in veins, is another example of difficult fluid-structure problems arising in blood flows. In addition, very large displacements and changes of topology (contact problems) have to be handled in those cases.

Due to stability reasons, it seems impossible to successfully apply in hemodynamics the explicit coupling schemes used in other fluid-structure problems, like aeroelasticity. As a result, fluid-structure interaction in biological flows raise new challenging issues in scientific computing and numerical analysis : new schemes have to be developed and analyzed.

We have proposed and analyzed over the last few years several efficient fluid-structure interaction algorithms. This topic remains very active. We are now using these algorithms to address inverse problems in blood flows to make patient specific simulations (for example, estimation of artery wall stiffness from medical imaging).

### *3.1.2. Aerosol*

Complex two-phase fluids can be modeled in many different ways. Eulerian models describe both phases by physical quantities such as the density, velocity or energy of each phase. In the mixed fluid-kinetic models, the biphasic fluid has one dispersed phase, which is constituted by a spray of droplets, with a possibly variable size, and a continuous classical fluid.

This type of model was first introduced by Williams [46] in the frame of combustion. It was later used to develop the Kiva code [36] at the Los Alamos National Laboratory, or the Hesione code [41], for example. It has a wide range of applications, besides the nuclear setting: diesel engines, rocket engines [39], therapeutic sprays, *etc.* One of the interests of such a model is that various phenomena on the droplets can be taken into account with an accurate precision: collision, breakups, coagulation, vaporization, chemical reactions, *etc.*, at the level of the droplets.

The model usually consists in coupling a kinetic equation, that describes the spray through a probability density function, and classical fluid equations (typically Navier-Stokes). The numerical solution of this system relies on the coupling of a method for the fluid equations (for instance, a finite volume method) with a method fitted to the spray (particle method, Monte Carlo).

We are mainly interested in modeling therapeutic sprays either for local or general treatments. The study of the underlying kinetic equations should lead us to a global model of the ambient fluid and the droplets, with some mathematical significance. Well-chosen numerical methods can give some tracks on the solutions behavior and help to fit the physical parameters which appear in the models.

## 3.2. Multiscale modeling

Multiscale modeling is a necessary step for blood and respiratory flows. In this section, we focus on blood flows. Nevertheless, similar investigations are currently carried out on respiratory flows.

### *3.2.1. Arterial tree modeling*

Problems arising in the numerical modeling of the human cardiovascular system often require an accurate description of the flow in a specific sensible subregion (carotid bifurcation, stented artery, *etc.*). The description of such local phenomena is better addressed by means of three-dimensional (3D) simulations, based on the numerical approximation of the incompressible Navier-Stokes equations, possibly accounting for compliant (moving) boundaries. These simulations require the specification of boundary data on artificial boundaries that have to be introduced to delimit the vascular district under study. The definition of such boundary conditions is critical and, in fact, influenced by the global systemic dynamics. Whenever the boundary data is not available from accurate measurements, a proper boundary condition requires a mathematical description of the action of the reminder of the circulatory system on the local district. From the computational point of view, it is not affordable to describe the whole circulatory system keeping the same level of detail. Therefore, this mathematical description relies on simpler models, leading to the concept of *geometrical multiscale* modeling of the circulation [42]. The underlying idea consists in coupling different models (3D, 1D or 0D) with a decreasing level of accuracy, which is compensated by their decreasing level of computational complexity.

The research on this topic aims at providing a correct methodology and a mathematical and numerical framework for the simulation of blood flow in the whole cardiovascular system by means of a geometric multiscale approach. In particular, one of the main issues will be the definition of stable coupling strategies between 3D and reduced order models.

To model the arterial tree, a standard way consists of imposing a pressure or a flow rate at the inlet of the aorta, *i.e.* at the network entry. This strategy does not allow to describe important features as the overload in the heart caused by backward traveling waves. Indeed imposing a boundary condition at the beginning of the aorta artificially disturbs physiological pressure waves going from the arterial tree to the heart. The only way to catch this physiological behavior is to couple the arteries with a model of heart, or at least a model of left ventricle.

A constitutive law for the myocardium, controlled by an electrical command, has been developed in the CardioSense3D project [0]. One of our objectives is to couple artery models with this heart model.

A long term goal is to achieve 3D simulations of a system including heart and arteries. One of the difficulties of this very challenging task is to model the cardiac valves. To this purpose, we investigate a mix of arbitrary Lagrangian Eulerian and fictitious domain approaches or x-fem strategies, or simplified valve models based on an immersed surface strategy.

### 3.2.2. Heart perfusion modeling

The heart is the organ that regulates, through its periodical contraction, the distribution of oxygenated blood in human vessels in order to nourish the different parts of the body. The heart needs its own supply of blood to work. The coronary arteries are the vessels that accomplish this task. The phenomenon by which blood reaches myocardial heart tissue starting from the blood vessels is called in medicine perfusion. The analysis of heart perfusion is an interesting and challenging problem. Our aim is to perform a three-dimensional dynamical numerical simulation of perfusion in the beating heart, in order to better understand the phenomena linked to perfusion. In particular the role of the ventricle contraction on the perfusion of the heart is investigated as well as the influence of blood on the solid mechanics of the ventricle. Heart perfusion in fact implies the interaction between heart muscle and blood vessels, in a sponge-like material that contracts at every heartbeat via the myocardium fibers.
Despite recent advances on the anatomical description and measurements of the coronary tree and on the corresponding physiological, physical and numerical modeling aspects, the complete modeling and simulation of blood flows inside the large and the many small vessels feeding the heart is still out of reach. Therefore, in order to model blood perfusion in the cardiac tissue, we must limit the description of the detailed flows at a given space scale, and simplify the modeling of the smaller scale flows by aggregating these phenomena into macroscopic quantities, by some kind of "homogenization" procedure. To that purpose, the modeling of the fluid-solid coupling within the framework of porous media appears appropriate.
Poromechanics is a simplified mixture theory where a complex fluid-structure interaction problem is replaced by a superposition of both components, each of them representing a fraction of the complete material at every point. It originally emerged in soils mechanics with the work of Terzaghi [45], and Biot [37] later gave a description of the mechanical behavior of a porous medium using an elastic formulation for the solid matrix, and Darcy's law for the fluid flow through the matrix. Finite strain poroelastic models have been proposed (see references in [38]), albeit with *ad hoc* formulations for which compatibility with thermodynamics laws and incompressibility conditions is not established.

### 3.2.3. Tumor and vascularization

The same way the myocardium needs to be perfused for the heart to beat, when it has reached a certain size, tumor tissue needs to be perfused by enough blood to grow. It thus triggers the creation of new blood vessels (angiogenesis) to continue to grow. The interaction of tumor and its micro-environment is an active field of research. One of the challenges is that phenomena (tumor cell proliferation and death, blood vessel adaptation, nutrient transport and diffusion, etc) occur at different scales. A multi-scale approach is thus being developed to tackle this issue. The long term objective is to predict the efficiency of drugs and optimize therapy of cancer.

[0] http://www-sop.inria.fr/CardioSense3D/

### *3.2.4. Respiratory tract modeling*

We aim at developing a multiscale model of the respiratory tract. Intraprenchymal airways distal from generation 7 of the tracheabronchial tree (TBT), which cannot be visualized by common medical imaging techniques, are modeled either by a single simple model or by a model set according to their order in TBT. The single model is based on straight pipe fully developed flow (Poiseuille flow in steady regimes) with given alveolar pressure at the end of each compartment. It will provide boundary conditions at the bronchial ends of 3D TBT reconstructed from imaging data. The model set includes three serial models. The generation down to the pulmonary lobule will be modeled by reduced basis elements. The lobular airways will be represented by a fractal homogenization approach. The alveoli, which are the gas exchange loci between blood and inhaled air, inflating during inspiration and deflating during expiration, will be described by multiphysics homogenization.

# 4. Application Domains

## 4.1. Blood flows

Cardiovascular diseases like atherosclerosis or aneurysms are a major cause of mortality. It is generally admitted that a better knowledge of local flow patterns could improve the treatment of these pathologies (although many other biophysical phenomena obviously take place in the development of such diseases). In particular, it has been known for years that the association of low wall shear stress and high oscillatory shear index give relevant indications to localize possible zones of atherosclerosis. It is also known that medical devices (graft or stent) perturb blood flows and may create local stresses favorable with atherogenesis. Numerical simulations of blood flows can give access to this local quantities and may therefore help to design new medical devices with less negative impacts. In the case of aneurysms, numerical simulations may help to predict possible zones of rupture and could therefore give a guide for treatment planning.

In clinical routine, many indices are used for diagnosis. For example, the size of a stenosis is estimated by a few measures of flow rate around the stenosis and by application of simple fluid mechanics rules. In some situations, for example in the case a sub-valvular stenosis, it is known that such indices often give false estimations. Numerical simulations may give indications to define new indices, simple enough to be used in clinical exams, but more precise than those currently used.

It is well-known that the arterial circulation and the heart (or more specifically the left ventricle) are strongly coupled. Modifications of arterial walls or blood flows may indeed affect the mechanical properties of the left ventricle. Numerical simulations of the arterial tree coupled to the heart model could shed light on this complex relationship.

One of the goals of the REO team is to provide various models and simulation tools of the cardiovascular system. The scaling of these models will be adapted to the application in mind: low resolution for modeling the global circulation, high resolution for modeling a small portion of vessel.

## 4.2. Respiratory tracts

Breathing, or “external” respiration (“internal” respiration corresponds to cellular respiration) involves gas transport though the respiratory tract with its visible ends, nose and mouth. Air streams then from the pharynx down to the trachea. Food and drink entry into the trachea is usually prevented by the larynx structure (epiglottis). The trachea extends from the neck into the thorax, where it divides into right and left main bronchi, which enter the corresponding lungs (the left being smaller to accommodate the heart). Inhaled air is then convected in the bronchus tree which ends in alveoli, where gaseous exchange occurs. Surfactant reduces the surface tension on the alveolus wall, allowing them to expand. Gaseous exchange relies on simple diffusion on a large surface area over a short path between the alveolus and the blood capillary under concentration gradients between alveolar air and blood. The lungs are divided into lobes (three on the right, two on the left) supplied by lobar bronchi. Each lobe of the lung is further divided into segments (ten segments of the right lung and eight of the left). Inhaled air contains dust and debris, which must be filtered, if possible, before they reach the alveoli. The tracheobronchial tree is lined by a layer of sticky mucus, secreted by the epithelium. Particles which hit the side wall of the tract are trapped in this mucus. Cilia on the epithelial cells move the mucous continually towards the nose and mouth.

Each lung is enclosed in a space bounded below by the diaphragm and laterally by the chest wall and the mediastinum. The air movement is achieved by alternately increasing and decreasing the chest pressure (and volume). When the airspace transmural pressure rises, air is sucked in. When it decreases, airspaces collapse and air is expelled. Each lung is surrounded by a pleural cavity, except at its hilum where the inner pleura give birth to the outer pleura. The pleural layers slide over each other. The tidal volume is nearly equal to 500 $ml$.

The lungs may fail to maintain an adequate supply of air. In premature infants surfactant is not yet active. Accidental inhalation of liquid or solid and airway infection may occur. Chronic obstructive lung diseases and lung cancers are frequent pathologies and among the three first death causes in France.

One of the goals of REO team in the ventilation field is to visualize the airways (virtual endoscopy) and simulate flow in image-based 3D models of the upper airways (nose, pharynx, larynx) and the first generations of the tracheobronchial tree (trachea is generation 0), whereas simple models of the small bronchi and alveoli are used (reduced-basis element method, fractal homogenization, multiphysics homogenization, lumped parameter models), in order to provide the flow distribution within the lung segments.

## 4.3. Cardiac electrophysiology

The purpose is to simulate the propagation of the action potential in the heart. A lot of works has already been devoted to this topic in the literature (see *e.g.* [40], [44], [43] and the references therein), nevertheless there are only very few studies showing realistic electrocardiograms obtained from partial differential equations models. Our goal is to find a compromise between two opposite requirements: on the one hand, we want to use predictive models, and therefore models based on physiology, on the other hand, we want to use models simple enough to be parametrized (in view of patient-specific simulations). One of the goal is to use our ECG simulator to address the inverse problem of electrocardiology. In collaboration with the Macs/M3disym project-team, we are interested in the electromechanical coupling in the myocardium. We are also interested in various clinical and industrial issues related to cardiac electrophysiology, in particular the simulation of experimental measurement of the field potential of cardiac stem cells in multi-electrode arrays.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

Chloé Audebert was awarded the AMIES PhD prize 2018 for her PhD thesis under the supervision of J.-F. Gerbeau and I. Vignon Clementel, in the framework of a collaboration with the SME company Fluoptics and with clinicians from Hôpital Paul Brousse (E. Vibert PUPH, Inserm 1193).

# 6. New Software and Platforms

## 6.1. FELiScE

*Finite Elements for Life SCiences and Engineering problems*
KEYWORDS: Finite element modelling - Cardiac Electrophysiology - Cardiovascular and respiratory systems

FUNCTIONAL DESCRIPTION: FELiScE is a finite element code which the M3DISIM and REO project-teams have decided to jointly develop in order to build up on their respective experiences concerning finite element simulations. One specific objective of this code is to provide in a unified software environment all the state-of-the-art tools needed to perform simulations of the complex respiratory and cardiovascular models considered in the two teams – namely involving fluid and solid mechanics, electrophysiology, and the various associated coupling phenomena. FELISCE is written in C++, and may be later released as an opensource library. FELiScE was registered in July 2014 at the Agence pour la Protection des Programmes under the Inter Deposit Digital Number IDDN.FR.001.350015.000.S.P.2014.000.10000.

- Participants: Axel Fourmont, Benoit Fabreges, Damiano Lombardi, Dominique Chapelle, Irène Vignon-Clementel, Jean-Frédéric Gerbeau, Marina Vidrascu, Matteo Aletti, Miguel Angel Fernandez Varela, Mikel Landajuela Larma, Philippe Moireau and Sébastien Gilles
- Contact: Miguel Angel Fernandez Varela
- URL: http://felisce.gforge.inria.fr

## 6.2. SHELDDON

*SHELls and structural Dynamics with DOmain decomposition in Nonlinear analysis*
FUNCTIONAL DESCRIPTION: SHELDDON is a finite element library based on the Modulef package which contains shell elements, nonlinear procedures and PVM subroutines used in domain decomposition or coupling methods, in particular fluid-structure interaction.

- Participants: Dominique Chapelle, Marina Vidrascu and Patrick Le Tallec
- Contact: Marina Vidrascu
- URL: https://gforge.inria.fr/projects/shelddon/

## 6.3. DCIMaL

KEYWORD: Cardiac Electrophysiology
FUNCTIONAL DESCRIPTION: DCIMaL is a Python and C++ software for safety pharmacology studies and particularly field potentials signals measured with micro-electrode array (MEA). The software includes a solver for field potential simulations and a dictionary of entries corresponding to features which can be extracted from real or simulated potential signals. It also includes an algorithm for drug classification (channel blockade or torsadogenic risk) and a tool for estimating ion channel activity (based on the CMAES library). DCIMaL was registered in 2018 at the Agence pour la Protection des Programmes Inter Deposit Digital Number IDDN.FR.001.270003.000.S.P.2018.000.31230

- Participants: Fabien Raphel, Jean-Frédéric Gerbeau and Damiano Lombardi
- Contact: Damiano Lombardi

## 6.4. FELiScE-NS

KEYWORDS: Incompressible flows - Thin-walled solids
FUNCTIONAL DESCRIPTION: FELiScE-NS is a set finite elements solvers for incompressible fluids (fractional-step schemes) and non-linear thin-walled structures (3D shells, and 2D curved beams) developed in the framework of the FELiScE library. FELiSCe-NS was registered in 2018 at the Agence pour la Protection des Programmes Inter Deposit Digital Number IDDN.FR.001.270015.000.S.A.2018.000.31200.

- Participants: Benoit Fabreges, Miguel Angel Fernandez Varela, Axel Fourmont, Jean-Frédéric Gerbeau and Marina Vidrascu
- Contact: Miguel Angel Fernandez Varela

# 7. New Results

## 7.1. Mathematical and numerical analysis of fluid-structure interaction problems

**Participants:** Muriel Boulakia, Ludovic Boilevin-Kayl, Chen-Yu Chiang, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Céline Grandmont, Damiano Lombardi, Marc Thiriet, Marina Vidrascu.

In [31], we consider a system modeling the interaction between a viscous incompressible fluid and an elastic structure. The fluid motion is represented by the classical Navier-Stokes equations while the elastic displacement is described by the linearized elasticity equation. The elastic structure is immersed in the fluid and the whole system is confined into a bounded domain of dimension 3. Our main result is the local in time existence and uniqueness of a strong solution of the corresponding system. This result holds without any restrictive assumptions on the domains geometry.

The numerical simulation of a thin-walled structure immersed in an incompressible fluid can be addressed by various methods. In [16], three of them are considered: the Arbitrary Lagrangian-Eulerian (ALE) method, the Fictitious Domain/Lagrange multipliers (FD) method and the Nitsche-XFEM method. Taking ALE as a reference, the advantages and limitations of FD and Nitsche-XFEM are carefully discussed on three benchmark test cases which have been chosen to be representative of typical difficulties encountered in valves or living cells simulations.

Fictitious domain approximations of fluid-structure interaction problems are generally discretized in time using strongly coupled schemes. This guarantees unconditional stability but at the price of solving a computationally demanding coupled system at each time-step. The design of loosely coupled schemes (i.e., methods that invoke the fluid and solid solvers only once per time-step) is of fundamental interest, especially for three-dimensional simulations, but the existing approaches are known to suffer from severe stability and/or time accuracy issues. In [28], we propose a new approach that overcomes these difficulties in the case of the coupling with immersed thin-walled structures.

In [27], we derive a Nitsche-based formulation for fluid-structure interaction (FSI) problems with contact. The approach is based on the work of Chouly and Hild [SIAM Journal on Numerical Analysis. 2013;51(2):1295-1307] for contact problems in solid mechanics. We present two numerical approaches, both of them formulating the FSI interface and the contact conditions simultaneously in equation form on a joint interface-contact surface. The first approach uses a relaxation of the contact conditions to allow for a small mesh-dependent gap between solid and wall. The second alternative introduces an artificial fluid below the contact surface. The resulting systems of equations can be included in a consistent fashion within a monolithic variational formulation, which prevents the so-called "chattering" phenomenon. To deal with the topology changes in the fluid domain at the time of impact, we use a fully Eulerian approach for the FSI problem. We compare the effect of slip and no-slip interface conditions and study the performance of the method by means of numerical examples.

## 7.2. Numerical methods for biological flows

**Participants:** Ludovic Boilevin-Kayl, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Florian Joly, Alexandre This, Marc Thiriet, Irene Vignon Clementel.

Cirrhosis is the common end-stage of chronic liver disease, with architectural distortion increasing the intrahepatic vascular resistance, leading to portal hypertension and systemic circulatory disorders. In [13] we investigate the impact of the changing vascular resistances on the hepatic and global circulation hemodynamics during cirrhogenesis. Morphological quantification of vascular trees from corosion casts of rats developing the disease provide the input for a lumped parameter model of the liver that was coupled to a model of the entire circulation of the rat. The simulations explain how vascular changes due to cirrhosis severely disrupt both hepatic and global hemodynamics.

Image-based models derived from CT angiography are being used clinically to simulate blood flow in the coronary arteries of individual patients to aid in the diagnosis of disease and planning treatments. However, image resolution limits vessel segmentation to larger epicardial arteries. In [20], we propose an algorithm for the generation of a patient-specific cardiac vascular network from epicardial vessels down to arterioles. We extend a tree generation method based on satisfaction of functional principles, to account for competing vascular trees, with flow-related and geometrical constraints adapting the simultaneous tree growths to patient priors.

Growth and remodeling of the embryo pharyngeal arch artery (PAA) network into the extracardiac great vessels is poorly understood but a major source of clinically serious malformations. In [21] we develop a methodological pipeline from high-resolution nano-computed tomography imaging and live-imaging flow measurements to multiscale pulsatile computational models. We identify local morphological variation along the PAAs and their association with specific hemodynamic changes in embryos of different stages, advancing our understanding of morphogenesis.

In [22] we evaluate atrioventricular valve regurgitation (AVVR) in babies born with an already very challenging heart condition, i.e., with single ventricle physiology. Although the second surgery that single ventricle patients undergo is thought to decrease AVVR, there is much controversy in the clinical literature about AVVR treatment. The effect of AVVR on Stage 1 haemodynamics and resulting acute changes from conversion to Stage 2 circulation in single ventricle patients are analyzed through lumped parameter models. Several degrees of AVVR severity are analyzed, for two types of valve regurgitation: incomplete leaflet closure and valve prolapse.

The medical imaging community is eager to define quantitative biophysical parameters. As part of a book addressing this question, in [26], we give a short overview of the mathematical modeling of blood flow at different resolutions, from the large vessel scale (three-dimensional, one-dimensional, and zero-dimensional modeling) to microcirculation and tissue perfusion.

In order to reduce the complexity of heart hemodynamics simulations, uncoupling approaches are often considered for the modeling of the immersed valves as an alternative to complex fluid-structure interaction (FSI) models. A possible shortcoming of these simplified approaches is the difficulty to correctly capture the pressure dynamics during the isovolumetric phases. In [35], we propose an enhanced resistive immersed surfaces (RIS) model of cardiac valves which overcomes this issue. The benefits of the model are investigated and tested in blood flow simulations of the left heart.

## 7.3. Numerical methods for cardiac electrophysiology

**Participants:** Muriel Boulakia, Jean-Frédéric Gerbeau, Damiano Lombardi, Fabien Raphel.

In [19] a method to assess the variability of phenomena described by PDEs is proposed. In particular, the probability density distribution of the parameters of a model is estimated, in such a way that the statistics of the model output match the observed ones. The investigated approach is based on a differential entropy regularised moment matching.

In [25] we investigated how, by a semi-empirical design of composite biomarkers, the classification of the action of a drug on the electrical activity of a cell can be improved. The data used are measured with a Micro-Electrodes-Array.

In [33] a method is investigated, to design composite biomarkers by exploiting a database of in silico experiments. In particular, a dictionary approach is proposed. The composite biomarker is expressed as a linear combination of linear and non-linear forms applied to the observable. The coefficients of the combination are determined by solving a $\ell^1$ regularised optimisation problem.

## 7.4. Lung and respiration modeling

**Participants:** Laurent Boudin, Céline Grandmont, Marina Vidrascu, Marc Thiriet, Irene Vignon Clementel.

In [34] we analyse multiscale models arising in the description of physiological flows such as blood flow in arteries or air flow in the bronchial tree. The fluid in the 3D part is described by the Stokes or the Navier-Stokes system which is coupled to 0D models or so-called Windkessel models. The resulting Navier-Stokes-Windkessel coupled system involves Neumann non-local boundary conditions that depends on the considered applications. We first show that the different types of Windkessel models share a similar formalism. Next we derive stability estimates for the continuous coupled Stokes-Windkessel or Navier-Stokes-Windkessel problem as well as stability estimates for the semi-discretized systems with either implicit or explicit coupling.We exhibit different kinds of behavior depending on the considered 0D model. Moreover even if no energy estimates can be derived in energy norms for the Navier-Stokes-Windkessel system, leading to possible numerical instabilities for large applied pressures, we show that stability estimates for both the continuous and semi-discrete problems, can be obtained in appropriate norms for small enough data by introducing a new well chosen Stokes-like operator. These sufficient stability conditions on the data may give a hint on the order of magnitude of the data enabling stable computations without stabilization method for the problem.

In [17], we consider a multi-species kinetic model which leads to the Maxwell-Stefan equations under a standard diffusive scaling (small Knudsen and Mach numbers). We propose a suitable numerical scheme which approximates both the solution of the kinetic model in rarefied regime and the one in the diffusion limit. We prove some a priori estimates (mass conservation and nonnegativity) and well-posedness of the discrete problem. We also present numerical examples where we observe the asymptotic-preserving behavior of the scheme.

In [30], we are interested in a system of fluid equations for mixtures with a stiff relaxation term of Maxwell-Stefan diffusion type. We use the formalism developed by Chen, Levermore, Liu to obtain a limit system of Fick type where the species velocities tend to align to a bulk velocity when the relaxation parameter remains small.

In [29], we consider the Boltzmann operator for mixtures with cutoff Maxwellian, hard potentials, or hard spheres collision kernels. In a perturbative regime around the global Maxwellian equilibrium, the linearized Boltzmann multi-species operator $L$ is known to possess an explicit spectral gap, in the global equilibrium weighted $L^2$ space. We study a new operator $L_\varepsilon$ obtained by linearizing the Boltzmann operator for mixtures around local Maxwellian distributions, where all the species evolve with different small macroscopic velocities of order $\varepsilon > 0$. This is a non-equilibrium state for the mixture. We establish a quasi-stability property for the Dirichlet form of $L_\varepsilon$ in the global equilibrium weighted $L^2$ space. More precisely, we consider the explicit upper bound that has been proved for the entropy production functional associated to $L$ and we show that the same estimate holds for the entropy production functional associated to $L_\varepsilon$, up to a correction of order $\varepsilon$.

## 7.5. Miscellaneous

**Participants:** Damiano Lombardi, Irene Vignon Clementel.

In [32] numerical quadrature schemes for the integration of observable quantities in the Brillouin zone for the periodic Schrödinger operator are investigated.

The indocyanine green (ICG) clearance, presented as plasma disappearance rate is, presently, a reliable method to estimate the hepatic function. However, this technique is not instantaneously available and thus cannot been used intra-operatively (during liver surgery). Near-infrared spectroscopy enables to assess hepatic ICG concentration over time in the liver tissue. In [14], we propose to extract more information from the liver intensity dynamics by interpreting it through a dedicated pharmacokinetics model. Parameters for different liver states are estimated from in-vivo measurements in rabbits (El-Desoky et al. 1999), and their link with liver function is investigated.

The hepatic hemodynamics is an essential parameter in surgical planning as well as in various disease processes. The transit time ultrasound (TTUS) perivascular flow probe technology is widely used in clinical practice to evaluate the hepatic inflow, yet invasive. The phase-contrast-MRI (PC-MRI) is not invasive and potentially applicable in assessing the hepatic blood flow. In [15], we compare the hepatic inflow rates using the PC-MRI and the TTUS probe, and evaluated their predictive value of post-hepatectomy adverse events in a porcine experimental model of partial hepatectomy.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

### *8.1.1. Philips Research*

**Participants:** Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Alexandre This.

CIFRE convention and contract with Philips Research for the PhD thesis of Alexandre This (January 2016 - December 2018) on fusion data/simulation for the assessment of mitral regurgitation.

### *8.1.2. Kephalios & Epygon*

**Participants:** Gautier Bureau, Miguel Ángel Fernández Varela, Jean-Frédéric Gerbeau, Ludovic Boilevin-Kayl, Marina Vidrascu.

REO is an academic partner of the industrial project MIVANA, dedicated to the development of new technologies for mitral valve treatment. It is led by the start-up company Kephalios, with the participation of the start-up company Epygon, by the company MDB Texinov and the research institute IFTH. In this framework, REO has two bilateral contracts with Kephalios and Epygon on the modeling and simulation of two medical devices for mitral valve repair.

### *8.1.3. Instem/NOTOCORD*

**Participants:** Muriel Boulakia, Damiano Lombardi, Jean-Frédéric Gerbeau, Fabien Raphel.

REO partners with the software company NOTOCORD. The collaboration started in 2013 the framework of the LabCom "cardioXcomp". In 2016, the ANR funding came to an end, and NOTOCORD was acquired by the company Instem. Our collaboration with Instem/NOTOCORD continues as a bilateral partnership with the purpose of developing the software cardioXcomp dedicated to the safety pharmacology industry. This project is also supported by a grant by AMIES (Agency for Interaction in Mathematics with Business and Society).

### *8.1.4. ESIEE-Heartflow*

**Participant:** Irene Vignon Clementel.

Research contract with ESIEE-Heartflow on coronary tree modeling.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR*

#### *9.1.1.1. ANR Project "IFSMACS"*

**Participants:** Muriel Boulakia, Céline Grandmont [local coordinator].

Period: 2015-2019.
The objective of this project, coordinated by Takéo Takahashi (Inria Nancy Grand-Est), is the mathematical analysis of systems involving structures immersed in a fluid. This includes the asymptotic analysis, the study of the controllability and stabilization of fluid-structure interaction systems, the understanding of the motion of self-propelled structures and the analysis and development of numerical methods to simulate fluid-structure systems.

*9.1.1.2. Participation to other ANR projects*

- Laurent Boudin is a member of the ANR Blanc project Kibord on kinetic models in biology and related domains
- Laurent Boudin is a member of the ANR TecSan Oxhelease
- Céline Grandmont is a member of the ANR TecSan Oxhelease
- Irene Vignon Clementel is a member of the project iLite (09/16-), RHU-santé grant, a large French hospital-medical research consortium that aims at developing innovations for liver and tissue engineering (Inria PI: Dirk Drasdo).

## 9.2. European Initiatives

### *9.2.1. Collaborations in European Programs, Except FP7 & H2020*

*9.2.1.1. SimInhale COST*

**Participant:** Irene Vignon Clementel.

Action MP1404, a pan-European network of experts in the field of inhaled medicine.

## 9.3. International Research Visitors

### *9.3.1. Visits of International Scientists*

*9.3.1.1. Internships*

- Charu Mittal, Visiting PhD student, Indian Institute of Technology Bombay, March 2018–August 2018

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. Member of the Organizing Committees*

- Laurent Boudin
  - Member of the organizing committee of the 7th "Forum Emploi Maths", December 2018, Paris
  - Member of both Boards of Mathematics Licence and Master, Sorbonne Université
  - Member of the IREM (Institutes for Research on Mathematics Teaching) Scientific Committee
  - Member of the SMAI (French Society for applied and industrial mathematics) Teaching Committee.
- Céline Grandmont
  - Co-organizer of Inria-LJLL meeting in scientific computing
  - Co-organizer of the CEMRACS 2018
- Irene Vignon Clementel
  - Minisymposium with F. Van de Vosse (Eindhoven U.) at WCCB, Dublin, Ireland.
  - Minisymposium at GDR mecabio January, Toulouse, France.
  - Session to foster collaboration between scientists and medical doctors at the Inria/CentraleSupelec/APHP meeting, November 12, Paris, France.

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Member of the Conference Program Committees*

- Irene Vignon Clementel
  - Steering committee, multidisciplinary workshop on surgical innovation WIC, 23-25 June, Cabourg, France
  - Programme committee member, Computational and Mathematical Biomedical Engineering Conference
  - Conference steering committee, International Conference on Engineering Frontiers in Pediatric and Congenital Heart Disease

*10.1.2.2. Member of the Editorial Boards*

- Jean-Frédéric Gerbeau
  - Member of the editorial board of the SIAM Journal of Scientific Computing (SISC).
  - Series editor of "SEMA SIMAI Series", Springer.
  - Member of the editorial board of Journal Advances in Computational Mathematics (ACOM), Springer
  - Member of the editorial board of International Journal for Numerical Methods in Biomedical Engineering (IJNMBE), Wiley.
  - Member of the editorial board of Communications in Applied and Industrial Mathematics, SIMAI/De Gruyter.
  - Member of the editorial board of Journal for Modeling in Ophthalmology, Kugler.
- Céline Grandmont
  - Member of the editorial board of Mathematical Modelling of Natural Phenomena
- Marc Thiriet
  - Member of the editorial board of Digital Medicine
  - Member of the editorial board of Computer Methods in Biomechanics and Biomedical Engineering–Imaging and Visualization
- Irene Vignon Clementel
  - Associate Editor of the International Journal for Numerical Methods in Biomedical Engineering

### *10.1.3. Leadership within the Scientific Community*

- Jean-Frédéric Gerbeau
  - Elected member of the Board of Directors of SMAI (French Society for Industrial and Applied Mathematics), in charge of the SMAI publications (M2AN, COCV, etc.)

### *10.1.4. Research Administration*

- Muriel Boulakia
  - Supervisor of the teaching of mathematics at the engineer school Polytech Sorbonne
- Miguel Ángel Fernández Varela
  - Deputy Head of Science, Inria Paris
  - Member of the Scientific Positions Commission, Inria Paris
  - Member of the Inria Evaluation Committee
- Jean-Frédéric Gerbeau
  - Head of science, Inria Paris (until Sept. 2017)
  - Member of the scientific committee of Labex NUMEV, Montpellier.

  - Service activity abroad: member of the Reference Committee of the PhD program Mathematical Models and Methods in Engineering (Politecnico di Milano, Italy).
- Céline Grandmont
  - Member of the Inria Evaluation Committee
  - Member of the Inria Parity Committee
- Irene Vignon-Clementel
  - Technology grant committee (Commission de développement technologique), Inria Paris center
  - Committee member for PhD students at Inria "Commission consultative des doctorants", since July 2016.
  - Mediator between PhD students and their supervisors for Inria Paris

### *10.1.5. Conferences*

- Ludovic Boilevin-Kayl
  - Contributed talk in minisymposium, 13th World Congress on Computational Mechanics, July 22th – 27th, 2018, New York City, USA.
  - Seminar, Jacques-Louis Lions Laboratory In-House Day, Sorbonne Université, April 5, 2018, Paris, France.
- Laurent Boudin
  - Seminar, Lab. de mathématiques appliquées du Havre, Univ. Le Havre Normandie, February 2018, Le Havre, France
  - Seminar, Cemracs 2018, August 2018, Marseille, France
  - Seminar of Partial Differential Equations, IRMA, Univ. Strasbourg, November 2018, Strasbourg, France
- Muriel Boulakia
  - congres Inverse Problems, Modeling and simulation, Malta, May 2018
  - Inria-LJLL meeting in scientific computing, October 2018
- Miguel Ángel Fernández Varela
  - Keynote talk in minisymposium, 13th World Congress on Computational Mechanics, July 22-27, 2018, New York City, USA.
- Felipe Galarce Marin
  - Contributed talk in minisymposium, 8th World Congress of Biomechanics, July 8-12, 2018, Dublin, Ireland
- Jean-Frédéric Gerbeau
  - Contributed talk in minisymposium, 8th World Congress of Biomechanics, July 8-12, 2018, Dublin, Ireland
  - Keynote talk in minisymposium, 6th European Conference on Computational Mechanics, June 11-15, 2018, Glasgow, UK
- Céline Grandmont
  - Seminar EDP, Nice, Jan. 2018
  - Seminar, Nantes, May 2018
  - Colloquium, Paris 5 Univ., Nov. 2018
  - Invited Speaker, Workshop Interfaces entre mathématiques et biologie, Nancy Univ., Nov. 2018
- Florian Joly

  - Weekend d'Innovation Chirugicale, Cabourg, Jun. 2018
  - Journées du GDR MécaBio à Montpellier, Nov. 2018
- Damiano Lombardi
  - CMM-Fields-Inria Workshop on Mathematics for Medicine, Toronto (Canada)
  - VPH 2018 Zaragoza (Spain), minisymposium of data assimilation
  - Workshop of matematics for biomedicine, Roma (Italy)
  - Ncardia workshop, Cologne (Germany)
- Marc Thiriet
  - Invited lecture, 15th International Symposium Computer Methods in Biomechanics and Biomedical Engineering and 3rd Conference on Imaging and Visualization March 27th – 30th, 2018, Lisbon, Portugal
  - Invited lecture, 2nd International Conference on Digital Medicine, May 25–27, 2018, Guangzhou, China
  - Invited lecture, France–Taiwan Science Festival, Sept. 14, 2018, Taipei, Taiwan
- Alexandre This
  - INdAM Workshop "Mathematical and Numerical Modeling of the Cardiovascular System",April 16 - 19, 2018, Rome, Italy.
  - Contributed talk in minisymposium, 8th World Congress of Biomechanics, July 8 - 12, 2018, Dublin, Ireland.
- Marina Vidrascu
  - Contributed talk in minisymposium, 13th World Congress on Computational Mechanics, July 22th – 27th, 2018, New York City, USA.
- Irene Vignon Clementel
  - Seminar (biomechanics), Ecole Polytechnique, Nov 16th, Palaiseau, France
  - Keynote lecture at the Workshop on Advanced Computational Biomechanics in Cardio-Vascular Surgery, Nov 8th, Saint-Etienne, France
  - Talk and poster, World conference of biomechanics, July 15-19 2018, Dublin, Ireland
  - Keynote, SBMC (systems biology in mammalian cells) conference, July 4-6, Brehmen, Germany
  - Invited talk, Weekend de l'innovation chirurgicale (WIC), Jun 22-24, Cabourg, France.
  - Invited talk, MRI and modeling workshop, June 22nd, Saclay, France
  - Poster presentation, EASL international conference (clinical liver conf.), April 11-15, Paris, France
  - Invited talk, GDR MecaBio (national biomechanics conference), Jan 10-12th, Toulouse, France

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Licence:

- Laurent Boudin
  - Introduction to series for signal theory, 18h, L2, UPMC
  - Calculus, 38.5h, L1, UPMC
  - Numerical methods for ODE, 53h, L3, UPMC
- Muriel Boulakia

  - Projects on differential equations, 34h, L3, Polytech Sorbonne, Sorbonne University
  - Nonlinear systems and optimization, 35h, L3, Polytech Sorbonne, Sorbonne University
  - Numerical approximation of functions, 36h, L3, Sorbonne University
- Jean-Frédéric Gerbeau
  - Control of dynamical systems, 32h, L3, Ecole Polytechnique.
- Damiano Lombardi
  - Analysis and Scientific Computing, 32h, L3, ENPC
  - Numerical Methods, 48h, L3, Polytech'Paris
  - Reduced models for problems in high dimension, 18h, ENPC, L3
- Irene Vignon Clementel
  - Numerical Methods for Ordinary Differential Equations, 24h ETD, L3, UPMC
  - Numerical simulations of blood flow, 2h30, as part of the undergraduate "continuum mechanics", AgroParisTech

Master:

- Laurent Boudin
  - Basics for numerical methods, 29h, M1, UPMC
  - Student advising for orientation and professional insertion, 20h, M1, UPMC
- Muriel Boulakia
  - Preparatory course for teaching admission examination "Agrégation", 40h, M2, Sorbonne University
- Miguel Ángel Fernández Varela
  - Modeling and numerical methods for hemodynamics, 30h, M2, UPMC
- Irene Vignon Clementel
  - MEC 550 - Biofluid Mechanics and Mass Transport, M2, 1h30, Ecole Polytechnique (engineering school), France
  - Annual ZCCE workshop, M2, 1h30, College of Engineering, Swansea University
  - Innovations thrapeutiques: du fondamental à l'appliqué, bioengineering module, M2, 1h, Paul Brousse Hospital, France
  - IFSBM (Institut de Formation Supérieure BioMédicale), M2, 1h30, Marie Lannelongue Hospital, France
  - M2 Sciences Chirurgicales de l'Université Paris Sud, 1h30, France

### *10.2.2. Supervision*

PhD in progress: Ludovic Boilevin-Kayl, Modeling of cardiac implantable devices, since February 2016. Supervisors: J.-F. Gerbeau & M.A. Fernández Varela

PhD in progress: Alexandre This, Fusion data/simulation for the assessment of mitral regurgitation, since January 2016. Supervisor: J.-F. Gerbeau

PhD in progress: Chen-Yu Chiang, Transport on biological systems and some applications, since February 2016. Supervisor: M. Thiriet

PhD in progress: Felipe Galarce, Enhancing hemodynamics measurements with mathematical modeling, since December 2017. Supervisors: J.-F. Gerbeau & D. Lombardi.

PhD in progress: Fannie Maria Gerosa, Immersed boundary methods for fluid-structure interaction with topological changes, since January 2018. Supervisor: M.A. Fernández Varela
PhD in progress: David Michel, Mathematical analysis of fluid-kinetic coupled models, since September 2018. Supervisors: L. Boudin & A. Moussa
PhD in progress: Nicolas Golse, Contributions of anatomical and hemodynamic modeling of the liver in the anticipation, realization and teaching of liver surgery, since November 2018. Supervisors: E Vibert and I. Vignon-Clementel.

### *10.2.3. Juries*

- Laurent Boudin
  - PhD committee: Nisrine Outada, Sorbonne Université & Caddi Ayad University, Gentien Marois, Insa Toulouse, Onera & CEA
  - HDR committee: Ayman Moussa, Sorbonne Université
- Muriel Boulakia
  - PhD committee: Pierre-Elliott Bécue, Inria Bordeaux Sud-Ouest; Guillaume Delay, Univ Toulouse (referee); Sourav Mitra, Univ. Toulouse; Josef Kolumban, Univ. Dauphine; Fabien Wahl, Inria Paris
  - Hiring committee: CDI researcher, Univ. Bordeaux
- Miguel Angel Fernández Varela
  - PhD committee: Rabii Mlika, INSA Lyon (president), Guillaume Delay, Universtié de Toulouse (member), Karol Cascavita, ENPC (member)
  - Hiring committee: Inria (DR2).
- Céline Grandmont
  - Member of the “agrégation” jury in mathematics.
  - Hiring committees: Inria Rennes (CR2), Inria CRCN.
  - PHD committee: J.-C. Casanova, Toulouse University (referee), J. Kolomban, Dauphine Univ. (president), S. Girel, Lyon Univ. (referee), C. Taing, Sorbonne Univ.
  - HDR committee: A. Moussa, Sorbonne Univ.
- Irene Vignon Clementel
  - Hiring committee: Nice University (MdC), Inria (CRCN), Inria (DR2, admission).
  - PHD committee: Clara Jaquet, ESIEE (member), Jules Dichamp, IMFT (referee) Noémie Boissier, UPMC (co-advisor) Mohamed Bekheit, U. Paris-Saclay (president).

## 10.3. Popularization

### *10.3.1. Interventions*

- Céline Grandmont
  - Conference “Métier”: Master Maths students, UPMC, Oct 2018
  - Conference at "Rendez-vous des Jeunes Mathématiciennes et Informaticiennes", Inria Oct. 2018
  - High school discussion on scientific career, 18th January, Blanche de Castille, Le Chesnay, France
  - High school conference, 9th January, Blanche de Castille, Le Chesnay, France

# 11. Bibliography

## Major publications by the team in recent years

[1] L. BOUDIN, L. DESVILLETTES, C. GRANDMONT, A. MOUSSA.*Global existence of solutions for the coupled Vlasov and Navier-Stokes equations*, in "Differential and integral equations", November 2009, vol. 22, n$^{o}$ 11-12, p. 1247-1271, https://hal.archives-ouvertes.fr/hal-00331895

[2] L. BOUDIN, B. GREC, F. SALVARANI.*A mathematical and numerical analysis of the Maxwell-Stefan diffusion equations*, in "Discrete and Continuous Dynamical Systems - Series B", 2012, vol. 17, n° 5, p. 1427-1440 [*DOI :* 10.3934/DCDSB.2012.17.1427], https://hal.archives-ouvertes.fr/hal-00490511

[3] M. BOULAKIA, S. CAZEAU, M. A. FERNÁNDEZ, J.-F. GERBEAU, N. ZEMZEMI.*Mathematical Modeling of Electrocardiograms: A Numerical Study*, in "Annals of Biomedical Engineering", 2010, vol. 38, n° 3, p. 1071-1097 [*DOI :* 10.1007/S10439-009-9873-0], https://hal.inria.fr/inria-00400490

[4] M. BOULAKIA, S. GUERRERO.*Regular solutions of a problem coupling a compressible fluid and an elastic structure*, in "Journal de Mathématiques Pures et Appliquées", 2010, vol. 94, n° 4, p. 341-365 [*DOI :* 10.1016/J.MATPUR.2010.04.002], https://hal.inria.fr/hal-00648710

[5] J. CHRISTOPHE, T. ISHIKAWA, N. MATSUKI, Y. IMAI, K. TAKASE, M. THIRIET, T. YAMAGUCHI.*Patient-specific morphological and blood flow analysis of pulmonary artery in the case of severe deformations of the lung due to pneumothorax*, in "Journal of Biomechanical Science and Engineering", 2010, vol. 5, n° 5, p. 485-498, https://hal.inria.fr/inria-00543090

[6] M. A. FERNÁNDEZ, J. MULLAERT, M. VIDRASCU.*Explicit Robin-Neumann schemes for the coupling of incompressible fluids with thin-walled structures*, in "Computer Methods in Applied Mechanics and Engineering", 2013, vol. 267, p. 566-593 [*DOI :* 10.1016/J.CMA.2013.09.020], https://hal.inria.fr/hal-00784903

[7] J.-F. GERBEAU, D. LOMBARDI.*Approximated Lax Pairs for the Reduced Order Integration of Nonlinear Evolution Equations*, in "Journal of Computational Physics", May 2014, vol. 265, p. 246-269 [*DOI :* 10.1016/J.JCP.2014.01.047], https://hal.inria.fr/hal-00933172

[8] C. GRANDMONT, M. HILLAIRET.*Existence of global strong solutions to a beam-fluid interaction system*, in "Archive for Rational Mechanics and Analysis", 2016 [*DOI :* 10.1007/S00205-015-0954-Y], https://hal.inria.fr/hal-01138736

[9] P. MOIREAU, C. BERTOGLIO, N. XIAO, C. A. FIGUEROA, C. TAYLOR, D. CHAPELLE, J.-F. GERBEAU.*Sequential identification of boundary support parameters in a fluid-structure vascular model using patient image data*, in "Biomechanics and Modeling in Mechanobiology", July 2012, vol. 12, n° 3, p. 475-496 [*DOI :* 10.1007/S10237-012-0418-3], https://hal.inria.fr/hal-00760703

[10] S. PANT, B. FABRÈGES, J.-F. GERBEAU, I. VIGNON-CLEMENTEL.*A methodological paradigm for patient-specific multi-scale CFD simulations: from clinical measurements to parameter estimates for individual analysis*, in "International Journal for Numerical Methods in Biomedical Engineering", December 2014, vol. 30, n° 12, p. 1614–1648 [*DOI :* 10.1002/CNM.2692], https://hal.inria.fr/hal-01093879

[11] I. VIGNON-CLEMENTEL, A. MARSDEN, J. FEINSTEIN.*A Primer on Computational Simulation in Congenital Heart Disease for the Clinician*, in "Progress in Pediatric Cardiology", 2010, vol. 30, n° 1-2, p. 3-13, Fondation Leducq [*DOI :* 10.1016/J.PPEDCARD.2010.09.002], https://hal.inria.fr/inria-00542957

## Publications of the year

### Articles in International Peer-Reviewed Journal

[12] E. ABBATE, M. BOULAKIA, Y. COUDIÈRE, J.-F. GERBEAU, P. ZITOUN, N. ZEMZEMI.*In silico assessment of the effects of various compounds in MEA/hiPSC-CM assays: Modelling and numerical simulations*, in

"Journal of Pharmacological and Toxicological Methods", 2018, vol. 89, p. 59-72, https://hal.inria.fr/hal-01562673

[13] C. AUDEBERT, G. PEETERS, P. SEGERS, W. LALEMAN, D. MONBALIU, H. KORF, J. TREBICKA, I. VIGNON-CLEMENTEL, C. DEBBAUT.*Closed-loop lumped parameter modelling of hemodynamics during cirrhogenesis in rats*, in "IEEE Transactions on Biomedical Engineering", 2018 [*DOI :* 10.1109/TBME.2018.2793948], https://hal.archives-ouvertes.fr/hal-01696050

[14] C. AUDEBERT, I. VIGNON-CLEMENTEL.*Model and methods to assess hepatic function from indocyanine green fluorescence dynamical measurements of liver tissue*, in "European Journal of Pharmaceutical Sciences", 2018 [*DOI :* 10.1016/J.EJPS.2018.01.008], https://hal.archives-ouvertes.fr/hal-01696064

[15] M. BEKHEIT, C. AUDEBERT, P. BUCUR, H. ADRIANSEN, E. BLED, M. WARTENBERG, I. VIGNON-CLEMENTEL, E. VIBERT.*Transit time ultrasound perivascular flow probe technology is superior to MR imaging on hepatic blood flow measurement in a porcine model*, in "Hepatobiliary & Pancreatic Diseases International", December 2018, vol. 17, n$^{o}$ 6, p. 538-545 [*DOI :* 10.1016/J.HBPD.2018.07.009], https://hal.inria.fr/hal-01954795

[16] L. BOILEVIN-KAYL, M. A. FERNÁNDEZ, J.-F. GERBEAU.*Numerical methods for immersed FSI with thin-walled structures*, in "Computers and Fluids", May 2018 [*DOI :* 10.1016/J.COMPFLUID.2018.05.024], https://hal.inria.fr/hal-01704575

[17] A. BONDESAN, L. BOUDIN, B. GREC.*A numerical scheme for a kinetic model for mixtures in the diffusive limit using the moment method*, in "Numerical Methods for Partial Differential Equations", 2018 [*DOI :* 10.1002/NUM.22345], https://hal.archives-ouvertes.fr/hal-01727725

[18] M. A. FERNÁNDEZ, M. LANDAJUELA.*Splitting schemes and unfitted mesh methods for the coupling of an incompressible fluid with a thin-walled structure*, in "IMA Journal of Numerical Analysis", December 2018, https://hal.inria.fr/hal-01309462

[19] J.-F. GERBEAU, D. LOMBARDI, E. TIXIER.*A moment-matching method to study the variability of phenomena described by partial differential equations*, in "SIAM Journal on Scientific Computing", May 2018, vol. 40, n$^{o}$ 3, https://hal.archives-ouvertes.fr/hal-01391254

[20] C. JAQUET, L. NAJMAN, H. TALBOT, L. GRADY, M. SCHAAP, B. SPAIN, H. J. KIM, I. VIGNON-CLEMENTEL, C. A. TAYLOR.*Generation of patient-specific cardiac vascular networks: a hybrid image-based and synthetic geometric model*, in "IEEE Transactions on Biomedical Engineering", 2018 [*DOI :* 10.1109/TBME.2018.2865667], https://hal.archives-ouvertes.fr/hal-01869264

[21] S. E. LINDSEY, J. T. BUTCHER, I. VIGNON-CLEMENTEL.*Cohort-based multiscale analysis of hemodynamic-driven growth and remodeling of the embryonic pharyngeal arch arteries*, in "Development (Cambridge, England)", October 2018, vol. 145, n$^{o}$ 20, dev162578 [*DOI :* 10.1242/DEV.162578], https://hal.inria.fr/hal-01954783

[22] S. PANT, C. CORSINI, C. BAKER, T.-Y. HSIA, G. PENNATI, I. VIGNON-CLEMENTEL.*A Lumped Parameter Model to Study Atrioventricular Valve Regurgitation in Stage 1 and Changes Across Stage 2 Surgery in Single Ventricle Patients*, in "IEEE Transactions on Biomedical Engineering", November 2018, vol. 65, n$^{o}$ 11, p. 2450-2458 [*DOI :* 10.1109/TBME.2018.2797999], https://hal.inria.fr/hal-01954784

[23] N. POZIN, S. MONTESANTOS, I. KATZ, M. PICHELIN, I. VIGNON-CLEMENTEL, C. GRANDMONT.*Predicted airway obstruction distribution based on dynamical lung ventilation data: a coupled modeling-machine learning methodology*, in "International Journal for Numerical Methods in Biomedical Engineering", 2018 [*DOI :* 10.1002/CNM.3108], https://hal.archives-ouvertes.fr/hal-01568065

[24] F. RAPHEL, M. BOULAKIA, N. ZEMZEMI, Y. COUDIÈRE, J.-M. GUILLON, P. ZITOUN, J.-F. GERBEAU.*Identification of ion currents components generating field potential recorded in MEA from hiPSC-CM*, in "IEEE Transactions on Biomedical Engineering", 2018, vol. 65, n$^{o}$ 6, p. 1311-1319 [*DOI :* 10.1109/TBME.2017.2748798], https://hal.archives-ouvertes.fr/hal-01570341

[25] E. TIXIER, F. RAPHEL, D. LOMBARDI, J.-F. GERBEAU.*Composite biomarkers derived from Micro-Electrode Array measurements and computer simulations improve the classification of drug-induced channel block*, in "Frontiers in Physiology", 2018, vol. 8, n$^{o}$ 1096, p. 1-30 [*DOI :* 10.3389/FPHYS.2017.01096], https://hal.archives-ouvertes.fr/hal-01570819

### Scientific Books (or Scientific Book chapters)

[26] A. CAIAZZO, I. VIGNON-CLEMENTEL.*Mathematical Modeling of Blood Flow in the Cardiovascular System*, in "Quantification of Biophysical Parameters in Medical Imaging", Springer International Publishing, February 2018, p. 45-70, https://hal.inria.fr/hal-01955520

### Research Reports

[27] E. BURMAN, M. A. FERNÁNDEZ, S. FREI.*A Nitsche-based formulation for fluid-structure interactions with contact*, Inria, May 2018, n$^{o}$ RR-9172, https://hal.inria.fr/hal-01784841

### Other Publications

[28] L. BOILEVIN-KAYL, M. A. FERNÁNDEZ, J.-F. GERBEAU.*A loosely coupled scheme for fictitious domain approximations of fluid-structure interaction problems with immersed thin-walled structures*, June 2018, working paper or preprint, https://hal.inria.fr/hal-01811290

[29] A. BONDESAN, L. BOUDIN, M. BRIANT, B. GREC.*Stability of the spectral gap for the Boltzmann multispecies operator linearized around non-equilibrium Maxwell distributions*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01924308

[30] L. BOUDIN, B. GREC, V. PAVAN.*Diffusion models for mixtures using a stiff dissipative hyperbolic formalism*, May 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01786824

[31] M. BOULAKIA, S. GUERRERO, T. TAKAHASHI.*Well-posedness for the coupling between a viscous incompressible fluid and an elastic structure*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01939464

[32] E. CANCÈS, V. EHRLACHER, D. GONTIER, A. LEVITT, D. LOMBARDI.*Numerical quadrature in the Brillouin zone for periodic Schrödinger operators*, May 2018, https://arxiv.org/abs/1805.07144 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01796582

[33] J.-F. GERBEAU, D. LOMBARDI, E. TIXIER.*How to choose biomarkers in view of parameter estimation*, June 2018, working paper or preprint, https://hal.inria.fr/hal-01811158

[34] C. GRANDMONT, S. MARTIN.*Continuous and semi-discrete stability estimates for 3d/0d coupled systems modelling airflows and blood flows*, September 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01734706

[35] A. THIS, L. BOILEVIN-KAYL, M. A. FERNÁNDEZ, J.-F. GERBEAU.*Augmented Resistive Immersed Surfaces valve model for the simulation of cardiac hemodynamics with isovolumetric phases*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01944798

## References in notes

[36] A. AMSDEN, P. O'ROURKE, T. BUTLER.*A computer program for chemically reactive flows with sprays*, Los Alamos National Laboratory, 1989, n$^{o}$ LA-11560-MS

[37] M. A. BIOT.*Theory of propagation of elastic waves in a fluid-saturated porous solid. II higher frequency range*, in "J. Acoust. Soc. Am.", 1956, vol. 28, p. 179–191

[38] D. CHAPELLE, J. SAINTE-MARIE, J.-F. GERBEAU, I. VIGNON-CLEMENTEL.*A poroelastic model valid in large strains with applications to perfusion in cardiac modeling*, in "Computational Mechanics", 2010, vol. 46, n$^{o}$ 1, p. 91-101 [*DOI :* 10.1007/s00466-009-0452-x]

[39] J. DUPAYS, Y. FABIGNON, P. VILLEDIEU, G. LAVERGNE, G. ESTIVALEZES.*Some aspects of two phase flows in solid propellant rocket motors, in Solid propellant chemistry, combustion and interior ballistics*, in "Progress in Astronautics and Aeronautics", V. YANG, T. BRILL, W. PEN (editors), Academic Press, 2000, vol. 185

[40] G. LINES, P. GROTTUM, A. TVEITO.*Modeling the electrical activity of the heart. A bidomain model of the ventricles embedded in a torso*, in "Comput. Visual. Sci.", 2003, vol. 5, p. 195-213

[41] R. MOTTE.*A numerical method for solving particle-fluid equations*, in "Trends in numerical and physical modeling for industrial multiphase flows", Cargèse, France, 2000

[42] A. QUARTERONI, S. RAGNI, A. VENEZIANI.*Coupling between lumped and distributed models for blood flow problems*, in "Comput. Visual Sci.", 2001, vol. 4, p. 111–124

[43] F. SACHSE.*Computational Cardiology: Modeling of Anatomy, Electrophysiology, and Mechanics*, Springer-Verlag, 2004

[44] J. SUNDNES, G. LINES, X. CAI, B. NIELSEN, K.-A. MARDAL, A. TVEITO.*Computing the electrical activity in the heart*, Springer-Verlag, 2006

[45] K. TERZAGHI.*Theoretical Soil Mechanics*, John Wiley and Sons, New-York, 1943

[46] F. WILLIAMS.*Combustion theory*, 2nd, Benjamin Cummings, 1985

# Project-Team RITS

## Robotics & Intelligent Transportation Systems

RESEARCH CENTER
**Paris**

THEME
**Robotics and Smart environments**

# Table of contents

# Project-Team RITS

*Creation of the Team: 2014 February 17, updated into Project-Team: 2015 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.5. - Complex systems
A1.5.1. - Systems of systems
A1.5.2. - Communicating systems
A2.3. - Embedded and cyber-physical systems
A3.4. - Machine learning and statistics
A3.4.1. - Supervised learning
A3.4.5. - Bayesian methods
A3.4.6. - Neural networks
A3.4.8. - Deep learning
A5.3. - Image processing and analysis
A5.3.4. - Registration
A5.4. - Computer vision
A5.4.1. - Object recognition
A5.4.4. - 3D and spatio-temporal reconstruction
A5.4.5. - Object tracking and motion analysis
A5.4.6. - Object localization
A5.5.1. - Geometrical modeling
A5.9. - Signal processing
A5.10. - Robotics
A5.10.2. - Perception
A5.10.3. - Planning
A5.10.4. - Robot control
A5.10.5. - Robot interaction (with the environment, humans, other robots)
A5.10.6. - Swarm robotics
A5.10.7. - Learning
A6. - Modeling, simulation and control
A6.1. - Methods in mathematical modeling
A6.2.3. - Probabilistic methods
A6.2.6. - Optimization
A6.4.1. - Deterministic control
A6.4.3. - Observability and Controlability
A6.4.4. - Stability and Stabilization
A8.6. - Information theory
A8.9. - Performance evaluation
A9.2. - Machine learning
A9.5. - Robotics
A9.7. - AI algorithmics

**Other Research Topics and Application Domains:**

B5.6. - Robotic systems
B6.6. - Embedded systems
B7.1.2. - Road traffic
B7.2. - Smart travel
B7.2.1. - Smart vehicles
B7.2.2. - Smart road
B9.5.6. - Data science

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Fawzi Nashashibi [Inria, Senior Researcher, Team leader, HDR]
Guy Fayolle [Inria, Emeritus]
Jean-Marc Lasgouttes [Inria, Researcher]
Gérard Le Lann [Inria, Emeritus]
Anne Verroust-Blondet [Inria, Researcher, HDR]

**External Collaborator**

Itheri Yahiaoui [Univ de Reims Champagne-Ardennes]

**Technical Staff**

Mohammad Abualhoul [Inria]
Zayed Alsayed [Inria]
Younes Bouchaala [Inria, until Jan 2018]
Pierre Bourre [Inria, from Oct 2018]
Raoul de Charette [Inria]
Mohamed Hadded [Inria, until Sep 2018]
Ilias Xydias [Inria, until Jul 2018]
Armand Yvet [Inria]
Carlos Flores [Inria, from Dec 2018]

**PhD Students**

Pierre de Beaucorps [Inria]
Carlos Flores [Inria, until Nov 2018]
Fernando Garrido [Inria, until Apr 2018]
Farouk Ghallabi [CIFRE Renault]
Maximilian Jaritz [CIFRE Valeo]
Imane Mahtout [CIFRE Renault]
Kaouther Messaoud [Inria]
Francisco Navas [Inria, until Jul 2018]
Dinh-Van Nguyen [Vietnamese grant, until Nov 2018]
Renaud Poncelet [ENS Rennes, from Sep 2018]
Danut-Ovidiu Pop [Inria]
Luis Roldao [CIFRE AKKA]

**Visiting Scientist**

Jean Francois Lalonde [Université Laval, Canada, Oct 2018]

**Administrative Assistant**

Chantal Chazelas [Inria]

# 2. Overall Objectives

## 2.1. Overall Objectives

The focus of the project-team is to develop the technologies linked to Intelligent Transportation Systems (ITS) with the objective to achieve sustainable mobility by the improvement of the safety, the efficiency of road transport according to the recent "Intelligent Vehicle Initiative" launched by the DG Information Society of the European Commission (for "Smarter, Cleaner, and Safer Transport"). More specifically, we want to develop, demonstrate and test some innovative technologies under the framework of LaRA, "La Route Automatisée [0]" which covers all the advanced driver assistance systems (ADAS) and the traffic management systems going all the way to fully automated vehicles.
These developments are all based on the sciences and technologies of information and communications (STIC) and have the objective to bring significant improvements in the road transport sector through incremental or breakthrough innovations. The project-team covers fundamental R&D work on key technologies, applied research to develop techniques that solve specific problems, and demonstrator activities to evaluate and disseminate the results.
The scientific approach is focused on the analysis and optimization of road transport systems through a double approach:

1. the control of individual road vehicles to improve locally their efficiency and safety,
2. the design and control of large transportation systems.

The first theme on vehicle control is broadly based on signal processing and data fusion in order to have a better machine understanding of the situation a vehicle may encounter, and on robotics techniques to control the vehicle in order to help (or replace) the driver to avoid accidents while improving the performance of the vehicle (speed, comfort, mileage, emissions, noise...). The theme also includes software techniques needed to develop applications in a real-time distributed and complex environment with extremely high safety standards. In addition, data must be exchanged between the vehicles; communication protocols have thus to be adapted to and optimized for vehicular networks characteristics (e.g. mobility, road safety requirements, heterogeneity, density), and communication needs (e.g. network latency, quality of service, network security, network access control).
The second theme on modeling and control of large transportation systems is also largely dependent on STIC. The objective, there, is to improve significantly the performance of the transportation system in terms of throughput but also in terms of safety, emissions, energy while minimizing nuisances. The approach is to act on demand management (e.g. through information, access control or road charging) as well as on the vehicles coordination. Communications technologies are essential to implement these controls and are an essential part of the R&D, in particular in the development of technologies for highly dynamic networks.

In order to address those issues simultaneously, RITS is organized into three research axes, each of which being driven by a separate sub-team. The first axis addresses the traditional problem of vehicle guidance and autonomous navigation. The second axis focuses on the large scale deployment and the traffic analysis and modeling. The third axis deals with the problem of telecommunications from two points of view:

- *Technical*: design certified architectures enabling safe vehicle-to-vehicle and vehicle-to-vehicle communications obeying to standards and norm;
- *Fundamental*, design and develop appropriate architectures capable of handling thorny problems of routing and geonetworking in highly dynamic vehicular networks and high speed vehicles.

Of course, these three research sub-teams interact to build intelligent cooperative mobility systems.

[0] LaRA is a Joint Research Unit (JRU) associating three French research teams: Inria's project-team RITS, Mines ParisTech's CAOR and LIVIC.

# 3. Research Program

## 3.1. Vehicle guidance and autonomous navigation

**Participants:** Mohammad Abualhoul, Zayed Alsayed, Pierre de Beaucorps, Younes Bouchaala, Pierre Bourre, Raoul de Charette, Carlos Flores, Maximilian Jaritz, Fernando Garrido, Farouk Ghallabi, Shirsendu Halder, Imane Mahtout, Kaouther Messaoud, Francisco Navas, Fawzi Nashashibi, Dinh-Van Nguyen, Renaud Poncelet, Danut-Ovidiu Pop, Luis Roldao, Anne Verroust-Blondet, Itheri Yahiaoui.

There are three basic ways to improve the safety of road vehicles and these ways are all of interest to the project-team. The first way is to assist the driver by giving him better information and warning. The second way is to take over the control of the vehicle in case of mistakes such as inattention or wrong command. The third way is to completely remove the driver from the control loop.

All three approaches rely on information processing. Only the last two involve the control of the vehicle with actions on the actuators, which are the engine power, the brakes and the steering. The research proposed by the project-team is focused on the following elements:

- perception of the environment,
- planning of the actions,
- real-time control.

### *3.1.1. Perception of the road environment*

**Participants:** Zayed Alsayed, Raoul de Charette, Maximilian Jaritz, Farouk Ghallabi, Shirsendu Halder, Kaouther Messaoud, Fawzi Nashashibi, Dinh-Van Nguyen, Danut-Ovidiu Pop, Luis Roldao, Anne Verroust-Blondet, Itheri Yahiaoui.

Either for driver assistance or for fully automated guided vehicle purposes, the first step of any robotic system is to perceive the environment in order to assess the situation around itself. Proprioceptive sensors (accelerometer, gyrometer,...) provide information about the vehicle by itself such as its velocity or lateral acceleration. On the other hand, exteroceptive sensors, such as video camera, laser or GPS devices, provide information about the environment surrounding the vehicle or its localization. Obviously, fusion of data with various other sensors is also a focus of the research.
The following topics are already validated or under development in our team:

- relative ego-localization with respect to the infrastructure, i.e. lateral positioning on the road can be obtained by mean of vision (lane markings) and the fusion with other devices (e.g. GPS);
- global ego-localization by considering GPS measurement and proprioceptive information, even in case of GPS outage;
- road detection by using lane marking detection and navigable free space;
- detection and localization of the surrounding obstacles (vehicles, pedestrians, animals, objects on roads, etc.) and determination of their behavior can be obtained by the fusion of vision, laser or radar based data processing;
- simultaneous localization and mapping as well as mobile object tracking using laser-based and stereovision-based (SLAMMOT) algorithms.

Scene understanding is a large perception problem. In this research axis we have decided to use only computer vision as cameras have evolved very quickly and can now provide much more precise sensing of the scene, and even depth information. Two types of hardware setups were used, namely: monocular vision or stereo vision to retrieve depth information which allow extracting geometry information.

We have initiated several works:

- estimation of the ego motion using monocular scene flow. Although in the state of the art most of the algorithms use a stereo setup, researches were conducted to estimate the ego-motion using a novel approach with a strong assumption.
- bad weather conditions evaluations. Most often all computer vision algorithms work under a transparent atmosphere assumption which assumption is incorrect in the case of bad weather (rain, snow, hail, fog, etc.). In these situations the light ray are disrupted by the particles in suspension, producing light attenuation, reflection, refraction that alter the image processing.
- deep learning for object recognition. New works are being initiated in our team to develop deep learning recognition in the context of heterogeneous data.
- deep learning for vehicle motion prediction.

### *3.1.2. Cooperative Multi-sensor data fusion*

**Participant:** Fawzi Nashashibi.

Since data are noisy, inaccurate and can also be unreliable or unsynchronized, the use of data fusion techniques is required in order to provide the most accurate situation assessment as possible to perform the perception task. RITS team worked a lot on this problem in the past, but is now focusing on collaborative perception approach. Indeed, the use of vehicle-to-vehicle or vehicle-to-infrastructure communications allows an improved on-board reasoning since the decision is made based on an extended perception.

As a direct consequence of the electronics broadly used for vehicular applications, communication technologies are now being adopted as well. In order to limit injuries and to share safety information, research in driving assistance system is now orientating toward the cooperative domain. Advanced Driver Assistance System (ADAS) and Cybercars applications are moving towards vehicle-infrastructure cooperation. In such scenario, information from vehicle based sensors, roadside based sensors and a priori knowledge is generally combined thanks to wireless communications to build a probabilistic spatio-temporal model of the environment. Depending on the accuracy of such model, very useful applications from driver warning to fully autonomous driving can be performed.

The Collaborative Perception Framework (CPF) is a combined hardware/software approach that permits to see remote information as its own information. Using this approach, a communicant entity can see another remote entity software objects as if it was local, and a sensor object, can see sensor data of others entities as its own sensor data. Last year we developed the basic hardware modules that ensure the well functioning of the embedded architecture including perception sensors, communication devices and processing tools.

Finally, since vehicle localization (ground vehicles) is an important task for intelligent vehicle systems, vehicle cooperation may bring benefits for this task. A new cooperative multi-vehicle localization method using split covariance intersection filter was developed during the year 2012, as well as a cooperative GPS data sharing method.

In the first method, each vehicle estimates its own position using a SLAM (Simultaneous Localization And Mapping) approach. In parallel, it estimates a decomposed group state, which is shared with neighboring vehicles; the estimate of the decomposed group state is updated with both the sensor data of the ego-vehicle and the estimates sent from other vehicles; the covariance intersection filter which yields consistent estimates even facing unknown degree of inter-estimate correlation has been used for data fusion.

In the second GPS data sharing method, a new collaborative localization method is proposed. On the assumption that the distance between two communicative vehicles can be calculated with a good precision, cooperative vehicle are considered as additional satellites into the user position calculation by using iterative methods. In order to limit divergence, some filtering process is proposed: Interacting Multiple Model (IMM) is used to guarantee a greater robustness in the user position estimation.

Accidents between vehicles and pedestrians (including cyclists) often result in fatality or at least serious injury for pedestrians, showing the need of technology to protect vulnerable road users. Vehicles are now equipped with many sensors in order to model their environment, to localize themselves, detect and classify obstacles, etc. They are also equipped with communication devices in order to share the information with other road users and the environment. The goal of this work is to develop a cooperative perception and communication system, which merges information coming from the communications device and obstacle detection module to improve the pedestrian detection, tracking, and hazard alarming.

Pedestrian detection is performed by using a perception architecture made of two sensors: a laser scanner and a CCD camera. The laser scanner provides a first hypothesis on the presence of a pedestrian-like obstacle while the camera performs the real classification of the obstacle in order to identify the pedestrian(s). This is a learning-based technique exploiting adaptive boosting (AdaBoost). Several classifiers were tested and learned in order to determine the best compromise between the nature and the number of classifiers and the accuracy of the classification.

### *3.1.3. Planning and executing vehicle actions*

**Participants:** Pierre de Beaucorps, Carlos Flores, Fernando Garrido, Imane Mahtout, Fawzi Nashashibi, Francisco Navas, Renaud Poncelet, Anne Verroust-Blondet.

From the understanding of the environment, thanks to augmented perception, we have either to warn the driver to help him in the control of his vehicle, or to take control in case of a driverless vehicle. In simple situations, the planning might also be quite simple, but in the most complex situations we want to explore, the planning must involve complex algorithms dealing with the trajectories of the vehicle and its surroundings (which might involve other vehicles and/or fixed or moving obstacles). In the case of fully automated vehicles, the perception will involve some map building of the environment and obstacles, and the planning will involve partial planning with periodical recomputation to reach the long term goal. In this case, with vehicle to vehicle communications, what we want to explore is the possibility to establish a negotiation protocol in order to coordinate nearby vehicles (what humans usually do by using driving rules, common sense and/or non verbal communication). Until now, we have been focusing on the generation of geometric trajectories as a result of a maneuver selection process using grid-based rating technique or fuzzy technique. For high speed vehicles, Partial Motion Planning techniques we tested, revealed their limitations because of the computational cost. The use of quintic polynomials we designed, allowed us to elaborate trajectories with different dynamics adapted to the driver profile. These trajectories have been implemented and validated in the JointSystem demonstrator of the German Aerospace Center (DLR) used in the European project HAVEit, as well as in RITS's electrical vehicle prototype used in the French project ABV. HAVEit was also the opportunity for RITS to take in charge the implementation of the Co-Pilot system which processes perception data in order to elaborate the high level command for the actuators. These trajectories were also validated on RITS's cybercars. However, for the low speed cybercars that have pre-defined itineraries and basic maneuvers, it was necessary to develop a more adapted planning and control system. Therefore, we have developed a nonlinear adaptive control for automated overtaking maneuver using quadratic polynomials and Lyapunov function candidate and taking into account the vehicles kinematics. For the global mobility systems we are developing, the control of the vehicles includes also advanced platooning, automated parking, automated docking, etc. For each functionality a dedicated control algorithm was designed (see publication of previous years). Today, RITS is also investigating the opportunity of fuzzy-based control for specific maneuvers. First results have been recently obtained for reference trajectories following in roundabouts and normal straight roads.

## 3.2. V2X Communications for cooperative ITS

**Participants:** Gérard Le Lann, Mohammad Abualhoul, Younes Bouchaala, Fawzi Nashashibi.

Wireless communications are expected to play an essential role in ensuring road safety, road efficiency, and driving comfort o. Road safety applications often require relatively short response time and reliable information exchange between neighboring vehicles and road-side units in any road density condition. Because of the performance of the existing radio communications technology largely degrades with the increase of the

traffic density, the challenge of designing wireless communications solution suitable for safety applications is enabling reliable communications in highly dense scenarios.

To investigate this open problem and trade-off situations, RITS has been working on medium access control design for the IEEE 802.11p radio communication and the deployment of supportive solutions such as visible light communications and testing the use-cases for extreme traffic conditions and highly dense scenarios. The works have been carried out considering the vehicle behavior such as autonomous and connected vehicles merging, sharing, and convoy forming as platoon scenarios with considering the hard-safety requirements.

Unlike many of the road safety applications, the applications regarding road efficiency and comfort of road users, often require connectivity to the Internet. Based on our expertise in both Internet-based communications in the mobility context and in ITS, we are investigating the use of IPv6 (Internet Protocol version 6 which is going to replace the current version, IPv4, IoT) for vehicular communications, in a combined architecture supporting both V2V and V2I.

Communication contributions at RITS team have been working on channel modeling for both radio and visible light communications, and design of communications mechanisms, especially for security, service discovery, multicast, and Geo-Cast message delivery, and access point selection.

RITS-team has one of the latest certified standard communication hardware and tools supported by the partnership with the YoGoKo Company. All platforms (connected and autonomous vehicles) are equipped with state-of-art communication units On-Board-Units (OBU), where the Rocquencourt site equipped with two stationary Road-Side-Units (RSU) enabling all kind of tests and projects requirements

Below follows a more detailed description of the related research issues.

### *3.2.1. Visible light and radio communication for cooperative autonomous driving*

**Participants:** Mohammad Abualhoul, Fawzi Nashashibi.

With the extensive development of the automobile industry and the popularity of using personal road vehicles in the last decade, both traffic accidents and road congestion levels have rapidly increased. Taking advantage of advanced wireless communications to enable C-ITS can improve both road fluidity and driver comfort. Ensuring the safety requirements has been the primary interest of the standardization societies dedicated to developing C-ITS applications, in particular with the expected significant demand for a broad range of applications targeting these strict safety requirements. RF communication technology deploying IEEE 802.11p standard for vehicular applications have been dedicated to facilitating relatively medium communication range that supports high data rate for the vehicular environment, where the technology meant to operate within the road safety requirements level.

As a consequence of the accelerated increase of the wireless-based communication devices numbers for ITS applications, the RF communication solutions are pushed toward an insatiable demand for wireless networks data access and a remarkable increase in both latency and channel congestion levels. This instability introduced more usage constraints when C-ITS is required. An example of such applications where the safety requirements and usage constraints might be strictly sharp are the convoy-based ITS applications.

This research effort contributes to the autonomous vehicular communication and urban mobility improvements. The work addresses the main radio-based V2V communication limitations and challenges for ITS hard-safety applications and intends to deploy the vehicular lighting system as a supportive communication solution for convoy-based applications as an IVC [0]-enabled autonomous vehicle. The ultimate objectives of thisresearch was to implement, validate and integrate the VLC system within the existing C-ITS architecture by developing a VLC prototype, together with sufficient hand-over algorithms enabling VLC, RF, and perception-based solutions to ensure the maximum safety requirements and the continuous information exchange between vehicles. The feasibility and efficiency of the VLC-RF system implementation and hand-over algorithms were subjects to perform practical-based in-depth investigations of the system. In addition to the improvement in road capacity by utilizing the convoy-based autonomous driving systems.

[0]Inter Vehicle Communication

### 3.2.2. Regulation study for interoperability tests for cooperative driving

**Participants:** Mohammad Abualhoul, Fawzi Nashashibi.

The technological advances of autonomous and connected road vehicles have been shown an accelerating pace in the recent years. On the other hand, the regulations for autonomous, or driverless, road vehicles across Europe still deserve much attention and discussion

Therefore, RITS-Inria team plays a key element in one of the European demonstration-based projects (AUTOC-ITS), which aims to contribute to the regulation study for interoperability in the adoption of autonomous driving in European urban nodes. The regulation study done by RITS team and project partners meant to conduct a deployment of Cooperative Intelligent Transport Systems (C-ITS) in Europe by enhancing interoperability for autonomous vehicles [18]. The project activities and RITS contributions will also boost the role of C-ITS as the primary catalyst for any future implementation of autonomous driving scenarios in Europe. The final demonstration of different European partners will require the implementation and preparations of three pilots sites in three major European cities: Paris, Madrid, and Lisbon. Pilot locations in these major cities are chosen to be located along the European Atlantic Corridor for interoperability evaluation.

RITS-Inria is coordinating the French contribution by evaluating the deployment of C-ITS services in the A13-Paris, which belongs to the French part of the Atlantic Corridor.

Team Core contributions:

- Provide up to date feedback to contribute to the present EU and international regulations on autonomous vehicles.
- Build and evaluate the pilots experimentally by deploying fully autonomous vehicles and a Cooperative Intelligent Transport Systems (C-ITS).
- Define and evaluate a safety autonomous driving services, such as:
  - Roadworks warning.
  - Weather conditions.
  - Other hazardous notifications.
- Define and perform communication interoperability tests between deferent partners for different scenarios, messaging and hardware to ensure the compatibility in using the IEEE 802.11p standard.
- Study the extension of the results on large-scale deployment in other European countries.
- Contribute to the European standards organizations such as C-Roads, C-ITS platforms.

AUTOC-ITS project brings the road authorities from France, Span, and Portugal (DGT, ANSR, SANEF) and C-ITS experts from research institutes and universities (Inria, INDRA, UPM, UC, IPN) to carry out a cooperative work and contributes to the C-ITS Platform by bringing answers to the field of automation driving.

### 3.2.3. V2X radio communications for road safety applications

**Participants:** Mohammad Abualhoul, Fawzi Nashashibi.

The development work and generating proper components to facilitate communication requirements and to be deployed in different projects scenarios is one of the main ongoing activities by all RITS team members.

There are continuous activities on both theoretical modeling and experimental evaluation of the radio channel characteristics in vehicular networks, especially the radio quality, channel congestion, load allocations, congestion, and bandwidth availability.

Based on our previous expertise and studies, we develop mechanisms for efficient and reliable V2X communications, access point selection, handover algorithms which are especially dedicated to road safety and autonomous driving applications.

### *3.2.4. Safety-critical communications in intelligent vehicular networks*

**Participant:** Gérard Le Lann.

Intelligent vehicular networks (IVNs) are constituents of ITS. IVNs range from platoons with a lead vehicle piloted by a human driver to fully ad-hoc vehicular networks, a.k.a. VANETs, comprising autonomous/automated vehicles. Safety issues in IVNs appear to be the least studied in the ITS domain. The focus of our work is on safety-critical (SC) scenarios, where accidents and fatalities inevitably occur when such scenarios are not handled correctly. In addition to on-board robotics, inter-vehicular radio communications have been considered for achieving safety properties. Since both technologies have known intrinsic limitations (in addition to possibly experiencing temporary or permanent failures), using them redundantly is mandatory for meeting safety regulations. Redundancy is a fundamental design principle in every SC cyber-physical domain, such as, e.g., air transportation. (Optics-based inter-vehicular communications may also be part of such redundant constructs.) The focus of our on-going work is on safety-critical (SC) communications. We consider IVNs on main roads and highways, which are settings where velocities can be very high, thus exacerbating safety problems acceptable delays in the cyber space, and response times in the physical space, shall be very small. Human lives being at stake, such delays and response times must have strict (non-stochastic) upper bounds under worst-case conditions (vehicular density, concurrency and failures). Consequently, we are led to look for deterministic solutions.

**Rationale**

In the current ITS literature, the term *safety* is used without being given a precise definition. That must be corrected. In our case, a fundamental open question is: what is the exact meaning of *SC communications*? We have devised a definition, referred to as space-time bounds acceptability (STBA) requirements. For any given problem related to SC communications, those STBA requirements serve as yardsticks for distinguishing acceptable solutions from unacceptable ones with respect to safety. In conformance with the above, STBA requirements rest on the following worst-case upper bounds: $\lambda$ for channel access delays, and $\Delta$ for distributed inter-vehicular coordination (message dissemination, distributed agreement).

Via discussions with foreign colleagues, notably those active in the IEEE 802 Committee, we have comforted our early diagnosis regarding existing standards for V2V/V2I/V2X communications, such as IEEE 802.11p and ETSI ITS-G5: they are totally inappropriate regarding SC communications. A major flaw is the choice of CSMA/CA as the MAC-level protocol. Obviously, there cannot be such bounds as $\lambda$ and $\Delta$ with CSMA/CA. Another flaw is the choice of medium-range omnidirectional communications, radio range in the order of 250 m, and interference range in the order of 400 m. Stochastic delays achievable with existing standards are just unacceptable in moderate/worst-case contention conditions. Consider the following setting, not uncommon in many countries: a highway, 3 lanes each direction, dense traffic, i.e. 1 vehicle per 12.5 m. A simple calculation leads to the following result: any vehicle may experience (destructive) interferences from up to 384 vehicles. Even if one assumes some reasonable communications activity ratio, say 25%, one finds that up to 96 vehicles may be contending for channel access. Under such conditions, MAC-level delays and string-wide dissemination/agreement delays achieved by current standards fail to meet the STBA requirements by huge margins.

Reliance on V2I communications via terrestrial infrastructures and nodes, such as road-side units or WiFi hotspots, rather than direct V2V communications, can only lead to poorer results. First, reachability is not guaranteed: hazardous conditions may develop anywhere anytime, far away from a terrestrial node. Second, mixing SC communications and ordinary communications within terrestrial nodes is a violation of the very fundamental segregation principle: SC communications and processing shall be isolated from ordinary communications and processing. Third, security: it is very easy to jam or to spy on a terrestrial node; moreover, terrestrial nodes may be used for launching all sorts of attacks, man-in-the-middle attacks for example. Fourth, delays can only get worse than with direct V2V communications, since transiting via a node inevitably introduces additional latencies. Fifth, the delivery of every SC message must be acknowledged, which exacerbates the latency problems. Sixth, availability: what happens when a terrestrial node fails?

Trying to tweak existing standards for achieving SC communications is vain. That is also unjustified. Clearly, medium-range omnidirectional communications are unjustified for the handling of SC scenarios. By definition, accidents can only involve vehicles that are very close to each other. Therefore, short-range directional communications suffice. The obvious conclusion is that novel protocols and inter-vehicular coordination algorithms based on short-range direct V2V communications are needed. It is mandatory to check whether these novel solutions meet the STBA requirements. Future standards specifically aimed at SC communications in IVNs may emerge from such solutions.

**Naming and privacy**

Additionally, we are exploring the (re)naming problem as it arises in IVNs. Source and destination names appear in messages exchanged among vehicles. Most often, names are IP addresses or MAC addresses (plate numbers shall not be used for privacy reasons). A vehicle which intends to communicate with some vehicle, denoted $V$ here, must know which name *name(V)* to use in order to reach/designate $V$. Existing solutions are based on multicasting/broadcasting existential messages, whereby every vehicle publicizes its existence (name and geolocation), either upon request (replying to a Geocast) or spontaneously (periodic beaconing). These solutions have severe drawbacks. First, they contribute to overloading communication channels (leading to unacceptably high worst-case delays). Second, they amount to breaching privacy voluntarily. Why should vehicles reveal their existence and their time dependent geolocations, making tracing and spying much easier? Novel solutions are needed. They shall be such that:

- At any time, a vehicle can assign itself a name that is unique within a geographical zone centered on that vehicle (no third-party involved),
- No linkage may exist between a name and those identifiers (plate numbers, IP/MAC addresses, etc.) proper to a vehicle,
- Different (unique) names can be computed at different times by a vehicle (names can be short-lived or long-lived),
- *name(V)* at UTC time $t$ is revealed only to those vehicles sufficiently close to $V$ at time $t$, notably those which may collide with $V$.

We have solved the (re)naming problem in string/cohort formations [43]. Ranks (unique integers in any given string/cohort) are privacy-preserving names, easily computed by every member of a string, in the presence of string membership changes (new vehicles join in, members leave). That problem is open when considering arbitrary clusters of vehicles/strings encompassing multiple lanes.

## 3.3. Probabilistic modeling for large transportation systems

**Participants:** Mohamed Hadded, Guy Fayolle, Jean-Marc Lasgouttes, Ilias Xydias.

This activity concerns the modeling of random systems related to ITS, through the identification and development of solutions based on probabilistic methods and more specifically through the exploration of links between large random systems and statistical physics. Traffic modeling is a very fertile area of application for this approach, both for macroscopic (fleet management [41], traffic prediction) and for microscopic (movement of each vehicle, formation of traffic jams) analysis. When the size or volume of structures grows (leading to the so-called "thermodynamic limit"), we study the quantitative and qualitative (performance, speed, stability, phase transitions, complexity, etc.) features of the system.

In the recent years, several directions have been explored.

### *3.3.1. Traffic reconstruction*

Large random systems are a natural part of macroscopic studies of traffic, where several models from statistical physics can be fruitfully employed. One example is fleet management, where one main issue is to find optimal ways of reallocating unused vehicles: it has been shown that Coulombian potentials might be an efficient tool to drive the flow of vehicles. Another case deals with the prediction of traffic conditions, when the data comes from probe vehicles instead of static sensors.

While the widely-used macroscopic traffic flow models are well adapted to highway traffic, where the distance between junction is long (see for example the work done by the NeCS team in Grenoble), our focus is on a more urban situation, where the graphs are much denser. The approach we are advocating here is model-less, and based on statistical inference rather than fundamental diagrams of road segments. Using the Ising model or even a Gaussian Random Markov Field, together with the very popular Belief Propagation (BP) algorithm, we have been able to show how real-time data can be used for traffic prediction and reconstruction (in the space-time domain).

This new use of BP algorithm raises some theoretical questions about the ways the make the belief propagation algorithm more efficient:

- find the best way to inject real-valued data in an Ising model with binary variables [45];
- build macroscopic variables that measure the overall state of the underlying graph, in order to improve the local propagation of information [42];
- make the underlying model as sparse as possible, in order to improve BP convergence and quality [44].

### *3.3.2. Exclusion processes for road traffic modeling*

The focus here is on road traffic modeled as a granular flow, in order to analyze the features that can be explained by its random nature. This approach is complementary to macroscopic models of traffic flow (as done for example in the Opale team at Inria), which rely mainly on ODEs and PDEs to describe the traffic as a fluid.

One particular feature of road traffic that is of interest to us is the spontaneous formation of traffic jams. It is known that systems as simple as the Nagel-Schreckenberg model are able to describe traffic jams as an emergent phenomenon due to interaction between vehicles. However, even this simple model cannot be explicitly analyzed and therefore one has to resort to simulation.

One of the simplest solvable (but non trivial) probabilistic models for road traffic is the exclusion process. It lends itself to a number of extensions allowing to tackle some particular features of traffic flows: variable speed of particles, synchronized move of consecutive particles (platooning), use of geometries more complex than plain 1D (cross roads or even fully connected networks), formation and stability of vehicle clusters (vehicles that are close enough to establish an ad-hoc communication system), two-lane roads with overtaking.

The aspect that we have particularly studied is the possibility to let the speed of vehicle evolve with time. To this end, we consider models equivalent to a series of queues where the pair (service rate, number of customers) forms a random walk in the quarter plane $\mathbb{Z}_+^2$.

Having in mind a global project concerning the analysis of complex systems, we also focus on the interplay between discrete and continuous description: in some cases, this recurrent question can be addressed quite rigorously via probabilistic methods.

We have considered in [39] some classes of models dealing with the dynamics of discrete curves subjected to stochastic deformations. It turns out that the problems of interest can be set in terms of interacting exclusion processes, the ultimate goal being to derive hydrodynamic limits after proper scaling. A seemingly new method is proposed, which relies on the analysis of specific partial differential operators, involving variational calculus and functional integration. Starting from a detailed analysis of the Asymmetric Simple Exclusion Process (ASEP) system on the torus $\mathbb{Z}/n\mathbb{Z}$, the arguments a priori work in higher dimensions (ABC, multi-type exclusion processes, etc), leading to systems of coupled partial differential equations of Burgers' type.

### *3.3.3. Random walks in the quarter plane $\mathbb{Z}_+^2$*

This field remains one of the important *violon d'Ingres* in our research activities in stochastic processes, both from theoretical and applied points of view. In particular, it is a building block for models of many communication and transportation systems.

One essential question concerns the computation of stationary measures (when they exist). As for the answer, it has been given by original methods formerly developed in the team (see books and related bibliography). For instance, in the case of small steps (jumps of size one in the interior of $\mathbb{Z}_+^2$), the invariant measure $\{\pi_{i,j}, i,j \geq 0\}$ does satisfy the fundamental functional equation (see [2]):

$$Q(x,y)\pi(x,y) = q(x,y)\pi(x) + \widetilde{q}(x,y)\widetilde{\pi}(y) + \pi_0(x,y). \tag{1}$$

where the unknown generating functions $\pi(x,y), \pi(x), \widetilde{\pi}(y), \pi_0(x,y)$ are sought to be analytic in the region $\{(x,y) \in \mathbb{C}^2 : |x| < 1, |y| < 1\}$, and continuous on their respective boundaries.

The given function $Q(x,y) = \sum_{i,j} p_{i,j} x^i y^j - 1$, where the sum runs over the possible jumps of the walk inside $\mathbb{Z}_+^2$, is often referred to as the *kernel*. Then it has been shown that equation (1) can be solved by reduction to a boundary-value problem of Riemann-Hilbert type. This method has been the source of numerous and fruitful developments. Some recent and ongoing works have been dealing with the following matters.

- *Group of the random walk*. In several studies, it has been noticed that the so-called *group of the walk* governs the behavior of a number of quantities, in particular through its *order*, which is always even. In the case of small jumps, the algebraic curve $R$ defined by $\{Q(x,y) = 0\}$ is either of *genus* 0 (the sphere) or 1 (the torus). In [Fayolle-2011a], when the drift of the random walk is equal to 0 (and then so is the genus), an effective criterion gives the *order* of the group. More generally, it is also proved that whenever the genus is 0, this order is infinite, except precisely for the zero drift case, where finiteness is quite possible. When the *genus* is 1, the situation is more difficult. Recently [40], a criterion has been found in terms of a determinant of order 3 or 4, depending on the arity of the group.
- *Nature of the counting generating functions*. Enumeration of planar lattice walks is a classical topic in combinatorics. For a given set of allowed jumps (or steps), it is a matter of counting the number of paths starting from some point and ending at some arbitrary point in a given time, and possibly restricted to some regions of the plane. A first basic and natural question arises: how many such paths exist? A second question concerns the nature of the associated counting generating functions (CGF): are they rational, algebraic, holonomic (or D-finite, i.e. solution of a linear differential equation with polynomial coefficients)?

  Let $f(i,j,k)$ denote the number of paths in $\mathbb{Z}_+^2$ starting from $(0,0)$ and ending at $(i,j)$ at time $k$. Then the corresponding CGF

$$F(x,y,z) = \sum_{i,j,k \geq 0} f(i,j,k) x^i y^j z^k \tag{2}$$

  satisfies the functional equation

$$K(x,y)F(x,y,z) = c(x)F(x,0,z) + \widetilde{c}(y)F(0,y,z) + c_0(x,y), \tag{3}$$

  where $z$ is considered as a time-parameter. Clearly, equations (2) and (1) are of the same nature, and answers to the above questions have been given in [Fayolle-2010].
- *Some exact asymptotics in the counting of walks in $\mathbb{Z}_+^2$*. A new and uniform approach has been proposed about the following problem: *What is the asymptotic behavior, as their length goes to infinity, of the number of walks ending at some given point or domain (for instance one axis)?* The method in [Fayolle-2012] works for *both* finite or infinite groups, and for walks not necessarily restricted to excursions.

### *3.3.4. Simulation for urban mobility*

We have worked on various simulation tools to study and evaluate the performance of different transportation modes covering an entire urban area.

- Discrete event simulation for collective taxis, a public transportation system with a service quality comparable with that of conventional taxis.
- Discrete event simulation a system of self-service cars that can reconfigure themselves into shuttles, therefore creating a multimodal public transportation system; this second simulator is intended to become a generic tool for multimodal transportation.
- Joint microscopic simulation of mobility and communication, necessary for investigation of cooperative platoons performance.

These two programs use a technique allowing to run simulations in batch mode and analyze the dynamics of the system afterward.

# 4. Application Domains

## 4.1. Introduction

While the preceding section focused on methodology, in connection with automated guided vehicles, it should be stressed that the evolution of the problems which we deal with remains often guided by the technological developments. We enumerate three fields of application whose relative importance varies with time and which have strong mutual dependencies: driving assistance, cars available in self-service mode and fully automated vehicles (cybercars).

## 4.2. Driving assistance

Several techniques will soon help drivers. One of the first immediate goal is to improve security by alerting the driver when some potentially dangerous or dangerous situations arise, i.e. collision warning systems or lane tracking could help a bus driver and surrounding vehicle drivers to more efficiently operate their vehicles. Human factors issues could be addressed to control the driver workload based on additional information processing requirements. Another issue is to optimize individual journeys. This means developing software for calculating optimal (for the user or for the community) paths. Nowadays, path planning software is based on a static view of the traffic: efforts have to be done to take the dynamic component in account.

## 4.3. New transportation systems

The problems related to the abusive use of the individual car in large cities led the populations and the political leaders to support the development of public transport. A demand exists for a transport of people and goods which associates quality of service, environmental protection and access to the greatest number. Thus the tram and the light subways of VAL type recently introduced into several cities in France conquered the populations, in spite of high financial costs. However, these means of mass transportation are only possible on lines on which there is a keen demand. As soon as one moves away from these “lines of desire” or when one deviates from the rush hours, these modes become expensive and offer can thus only be limited in space and time. To give a more flexible offer, it is necessary to plan more individual modes which approach the car as we know it. However, if one wants to enjoy the benefits of the individual car without suffering from their disadvantages, it is necessary to try to match several criteria: availability anywhere and anytime to all, lower air and soils pollution as well as sound levels, reduced ground space occupation, security, low cost. Electric or gas vehicles available in self-service, as in the Praxitèle system, bring a first response to these criteria. To be able to still better meet the needs, it is however necessary to re-examine the design of the vehicles on the following points:

- ease empty car moves to better distribute them;
- better use of information systems inboard and on ground;
- better integrate this system in the global transportation system.

These systems are now operating. The challenge is to bring them to an industrial phase by transferring technologies to these still experimental projects.

## 4.4. Automated vehicles

The long term effort of the project is to put automatically guided vehicles (cybercars) on the road. It seems too early to mix cybercars and traditional vehicles, but data processing and automation now make it possible to consider in the relatively short term the development of such vehicles and the adapted infrastructures. RITS aims at using these technologies on experimental platforms (vehicles and infrastructures) to accelerate the technology transfer and to innovate in this field. Other application can be precision docking systems that will allow buses to be automatically maneuvered into a loading zone or maintenance area, allowing easier access for passengers, or more efficient maintenance operations. Transit operating costs will also be reduced through decreased maintenance costs and less damage to the braking and steering systems. Regarding technical topics, several aspects of Cybercars have been developed at RITS this year. First, we have stabilized a generic Cycab architecture involving Inria SynDEx tool and CAN communications. The critical part of the vehicle is using a real-time SynDEx application controlling the actuators via two Motorola's MPC555. Today, we have decided to migrate to the new dsPIC architecture for more efficiency and ease of use. This application has a second feature, it can receive commands from an external source (Asynchronously to this time) on a second CAN bus. This external source can be a PC or a dedicated CPU, we call it high level. To work on the high level, in the past years we have been developing a R&D framework called (Taxi) which used to take control of the vehicle (Cycab and Yamaha) and process data such as gyro, GPS, cameras, wireless communications and so on. Today, in order to rely on a professional and maintained solution, we have chosen to migrate to the RTMaps SDK development platform. Today, all our developments and demonstrations are using this efficient prototyping platform. Thanks to RTMaps we have been able to do all the demonstrations on our cybercars: cycabs, Yamaha AGV and new Cybus platforms. These demonstrations include: reliable SLAMMOT algorithm using 2 to 4 laser sensors simultaneously, automatic line/road following techniques, PDA remote control, multi sensors data fusion, collaborative perception via ad-hoc network. The second main topic is inter-vehicle communications using ad-hoc networks. We have worked with the EVA team for setting and tuning OLSR, a dynamic routing protocol for vehicles communications. Our goal is to develop a vehicle dedicated communication software suite, running on a specialized hardware. It can be linked also with the Taxi Framework for getting data such GPS information's to help the routing algorithm.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

Mohammad Abualhoul, with the paper , won the Runner-up Best Paper Award at ICVES 2018 (2018 IEEE International Conference on Vehicular Electronics and Safety, September 12-14, Madrid, Spain).
BEST PAPERS AWARDS :
[17]
M. ABUALHOUL, E. TALAVERA MUNOZ, F. NASHASHIBI.*The Use of Lane-Centering to Ensure the Visible Light Communication Connectivity for a Platoon of Autonomous Vehicles*, in "ICVES'2018 - 20th IEEE International Conference on Vehicular Electronics and Safety", Madrid, Spain, September 2018, https://hal.inria.fr/hal-01888549

# 6. New Software and Platforms

## 6.1. PML-SLAM

KEYWORD: Localization

SCIENTIFIC DESCRIPTION: Simultaneous Localization and Mapping method based on 2D laser data.

- Participants: Fawzi Nashashibi and Zayed Alsayed
- Contact: Fawzi Nashashibi

## 6.2. V2Provue

*Vehicle-to-Pedestrian*
FUNCTIONAL DESCRIPTION: It is a software developed for the Vehicle-to-Pedestrian (V2P) communications, risk calculation, and alarming pedestrians of collision risk. This software is made of an Android application dedicated to pedestrians and RtMaps modules for the vehicles.

On the pedestrian side, the application is relying on GPS data to localize the user and Wi-Fi communications are used to receive messages about close vehicles and send information about the pedestrian positioning. Besides, a service has been developed to evaluate the collision risk with the vehicles near the pedestrian and an HMI based on OpenStreetMap displays all the useful information such as pedestrian and vehicles localization and, collision risk.

On the vehicle side, RtMaps modules allowing V2X communications have been developed. These modules contain features such as TCP/UDP socket transmissions, broadcast, multicast, unicast communications, routing, forwarding algorithms, and application specific modules. In the V2ProVu software, a particular application module has been implemented to create data packets containing information about the vehicle state (position, speed, yaw rate,...) and the V2X communication stack is used to broadcast these packets towards pedestrians. Moreover, the V2proVu application can also receive data from pedestrians and create objects structures that can be shared with the vehicle perception tools.

- Contact: Fawzi Nashashibi

## 6.3. SimConVA

*Connected Autonomous Vehicles Simulator*
FUNCTIONAL DESCRIPTION: The software provides an interface between the network simulator ns-3 (https://www.nsnam.org/) and the modular prototyping framework RTMaps (https://intempora.com/).

This code allows to create an RTMaps component which activates and controls the ns-3 simulator. The component handles the sending and reception of data packets between ns-3 and RTMaps for each vehicle. It also handles the mobility of vehicles in ns-3 using their known position in RTMaps.

- Authors: Pierre Merdrignac, Oyunchimeg Shagdar and Jean-Marc Lasgouttes
- Contact: Jean-Marc Lasgouttes

# 7. New Results

## 7.1. Deep Reinforcement Learning for end-to-end driving

**Participants:** Raoul de Charette, Maximilian Jaritz, Fawzi Nashashibi.

Following the work initiated in 2017, we continued the work on end-to-end driving using with asynchronous reinforcement learning directly. The network learns to map low level control directly with RGB images. To continue previous works initiated, we have applied recent domain adaptation and evaluated our reinforcement learning (learn in a realistic car game) in open-loop on real video footage, showing promising adaptation results. New outcome also include tests on real data (web footage). This led to a publication in ICRA [25]. This research was partially funded by Valeo.

## 7.2. Convolutional neural networks for Semantic and Completion with Sparse and Dense Data

**Participants:** Raoul de Charette, Maximilian Jaritz, Fawzi Nashashibi.

Deep convolutional networks have outperform all previous techniques on most vision tasks. This is because they are able to utilize dense data and extract relationship between local information such as gradients, or high level features. However, convolutional neural networks (CNNs) require dense data and are known to fail when data is sparse. Here, we address the research problem and proposed a solution. Instead of using a sparse convolution methodology, we show that using the right architecture with a proper training strategy the network can learn sparsity invariant feature while remaining stable when dense data are present. Our architecture uses an encoder-decoder version of Mobile NasNet with skip connections. The results show that we can accomplish both data completion or semantic segmentation changing only the last layer of the network. Performance obtained were published on Kitti Benchmark and ranks among the first ones, and the methodology was published in 3DV [26]. This research was partially funded by Valeo.

## 7.3. Realistic Weather Augmentation for Evaluation of Bad Weather in Computer Vision

**Participants:** Raoul de Charette, Shirsendu Halder.

Computer vision is evaluated on extensive databases that include large number of examples and allow the ranking of algorithms. However, all databases are acquired in clear weather conditions, where the atmosphere is a transparent medium. In rain/snow/fog, when the atmosphere is filled with particles the light is refracted/reflected/diffracted and the appearance is altered. Here we propose a new research that augment existing databases with new weather or arbitrary amount. We applied it on Kitti and Cityscapes. Our approach uses an accurate understanding of physical and optics models to generate realistic rain/fog and augment existing images or sequences. This allows us to evaluate state-of-the-art vision algorithms for both object detection and semantics and quantitatively measure the effect of rain or fog on them. This research was conducted in collaboration with Jean-Francois Lalonde from Université Laval and was supported by Samuel de Champlain Quebec-France collaboration program.

## 7.4. Perception for Cooperative Driving

**Participants:** Pierre Bourre, Raoul de Charette, Carlos Flores, Renaud Poncelet, Luis Roldao, Dinh-Van Nguyen.

In the context of multiple autonomous vehicles, sharing the perception of each other allows an enriched perception of the environment. For the PACV2x FUI project, we propose a mix of vision sensors and communication exchanges is used for merging, overtaking, and other risky situations that benefit from multi perception. A speed planning algorithm as well as low level control and lidar data clustering were developed to allow a small fleet of two to three vehicles to handle such scenarios. The vehicles use communication and GPS coordinates to closely follow a planned trajectory.

## 7.5. A Statistical Update of Grid Representations from Range Sensors

**Participants:** Luis Roldao, Raoul de Charette, Anne Verroust-Blondet.

An accurate 3D model of the surrounding environment is a fundamental feature for autonomous vehicles to perform different tasks such as obstacle detection, localization and mapping. While continuous representations are widely used in the literature, we prefer to use a three dimensional discrete grid representation in this work in order to reduce memory an computational complexity. In this case, each grid cell represents the occupancy state of a portion of the environment in a probabilistic manner.

By definition, a discretized representation inhibits a completely accurate reconstruction. Therefore, grid models are unable to create a perfect model of the surroundings. In the literature, it is usually considered that within a single scan, the state of each cell is binary (free or occupied). Hence, a cell is set occupied if at least one impact occurred within, and free if it has been traversed by any ray. The problem of such an approach is that the complete state of the cell is updated from a single partial observation, neglecting the contribution of multiple measurements and their validity. Moreover, the traversed distance of the rays within each cell is usually ignored.

Towards the goal of achieving a more accurate representation, we propose a different way to update the occupancy probability of each cell according to the observations; considering the traversed distance of the rays within each cell (ray-path information), the contribution of the complete set of observations within the cell, and the density of observations that can be obtained at such cell according to its distance from the sensor. Proposed method was evaluated in both simulated and real data. Reconstruction results show an improvement on the representation of the surroundings with less occupancy state errors in the cells of the grid. Future works will include the comparison against a continuous representation to test the accuracy along with the time and computation needs for both representations.
More details can be found in [38] and [30]. This research is partially funded by AKKA Technology.

## 7.6. Recognizing Pedestrians using Cross-Modal Convolutional Networks

**Participants:** Danut-Ovidiu Pop, Fawzi Nashashibi.

This year, we have continued our research, which is based on multi-modal image fusion schemes with deep learning classification methods. We propose four different learning patterns based on Cross-Modality deep learning of Convolutional Neural Networks:

(1) a Particular Cross-Modality Learning;

(2) a Separate Cross-Modality Learning;

(3) a Correlated Cross-Modality Learning and

(4) an Incremental CrossModality Learning model.
Moreover, we also design a new variation of a Lenet architecture, which improves the classification performance. Finally, we propose to learn this model with the incremental cross-modality approach using optimal learning settings, obtained with a K-fold Cross Validation pattern. This method outperforms the state-of-the-art classifier provided with Daimler datasets on both non-occluded and partially-occluded pedestrian tasks.

## 7.7. Vehicle Trajectory Prediction

**Participants:** Kaouther Messaoud, Itheri Yahiaoui, Anne Verroust-Blondet, Fawzi Nashashibi.

In order to enhance the road safety, the first and the most important step for an effective autonomous navigation is the environment perception and surrounding objects recognition. So, advanced sensing systems are mounted in vehicles to monitor the on-road environment. One of the most challenging tasks is to understand, analyze the driving situations and make a reasonable and safe navigation decisions accordingly. Human drivers make decisions while implicitly reasoning about how neighboring drivers will move in the future. In this context, we aim to predict the motion of drivers neighboring an autonomous vehicle based on data captured using deployed sensors.
This year, we studied the state of the art approaches for trajectory and maneuver prediction. We focused on general trajectory prediction representation while considering interactions between the neighboring drivers using different types of neural networks such as recurrent and convolutional neural networks.

## 7.8. WiFi Fingerprinting Localization for Intelligent Vehicles in Car Park

**Participants:** Dinh-Van Nguyen, Raoul de Charette, Fawzi Nashashibi.

A novel method of WiFi fingerprinting for localizing intelligent vehicles in GPS-denied area, such as car parks, has been proposed. Although the method itself is a popular approach for indoor localization application, adapting it to the speed of vehicles requires different treatment. By deploying an ensemble neural network for fingerprinting classification, the method shows a reasonable localization precision at car park speed. Furthermore, a Gaussian Mixture Model (GMM) Particle Filter is applied to increase localization frequency as well as accuracy. Experiments show promising results with average localization error of 0.6m (cf. [29]).

A more complete study on the use of Wifi fingerprinting for solving the localization problem for autonomous vehicles in GPS-denied environments is presented in the thesis manuscript entitled "Wireless Sensors Networks for Indoor Mapping and Accurate Localization for Low Speed Navigation in Smart Cities" (cf. [11]).

## 7.9. Enhancing the Accuracy of SLAM-based Localization Systems for Autonomous Driving

**Participants:** Zayed Alsayed, Anne Verroust-Blondet, Fawzi Nashashibi.

Computing a reliable and accurate pose for a vehicle in any situation is one of the challenges for Simultaneous Localization And Mapping methods (SLAM) methods. Based on the probabilistic form of the SLAM solution, SLAM methods suffer from systematic errors related to the linearization of the solution models. The accuracy of the SLAM method can be improved by estimating a correction to be applied to the SLAM output based on relevant information available from the SLAM algorithm. In [20] two approaches predicting corrections to be applied to SLAM estimations are proposed:

1) The first approach is designed for 2D SLAM methods, i.e. independently of the underlying SLAM process and sensor used, where we aim to reduce the errors due to the dynamical modeling during specific maneuvers.

2) The second method is designed to handle errors related to the probabilistic formulation of Maximum Likelihood SLAM approaches, and thus it is suitable for 2D Maximum Likelihood SLAM methods (i.e. no assumptions on the sensor used).
The validity of both approaches was proved through two experiments using different evaluation metrics and using different sensor characteristics.
More detail can be found in the thesis manuscript of Zayed Alsayed entitled "Characterizing the Robustness and Enhancing the Accuracy of SLAM-based Localization Systems for Autonomous Driving" (cf. [7]).

## 7.10. LIDAR-based lane marking detection for vehicle localization

**Participants:** Farouk Ghallabi, Fawzi Nashashibi.

Accurate self-vehicle localization is an important task for autonomous driving and ADAS. Current GNSS-based solutions do not provide better than 2-3 m in open-sky environments. In order to achieve lane-level accuracy, a lane marking detection system using a multilayer LIDAR (velodyne) and a map matching algorithm has been introduced. The perception system includes three different steps: road segmentation, image construction and line detection. Our road segmentation method purely relies on geometric analysis of each layer returns. Detected lane markings are matched to a prototype third party map which was built with absolute accuracy = 5cm. The map matching algorithm is a particle filtering process that achieves lane-level accuracy (20 cm). More details are in [23]. This work has been partially funded by Renault.

## 7.11. Motion Planning among Highly Dynamic Obstacles

**Participants:** Pierre de Beaucorps, Anne Verroust-Blondet, Renaud Poncelet, Fawzi Nashashibi.

Motion planning in a dynamic environment is of great importance in many robotics applications. In the context of an autonomous mobile robot, it requires computing a collision-free path from a start to a goal among moving and static obstacles. We have introduced a framework to integrate into a motion planning method the interaction zones of a moving robot with its future surroundings, the reachable interaction sets (RIS). It can handle highly dynamic scenarios when combined with path planning methods optimized for quasi-static environments. It has been integrated with an artificial potential field reactive method and with a Bézier curve path planning. Experimental evaluations show that this approach significantly improves dynamic path planning methods, especially when the speeds of the obstacles are higher than the one of the robot (cf. [32] for more detail). This work has been partially funded by Valeo.

## 7.12. Control Architecture for Adaptive and Cooperative Car-Following

**Participants:** Carlos Flores, Fawzi Nashashibi.

The general scope of this work deals with three open challenges in the state-of-the-art of cooperative car-following systems:

1) Deal with the impact of not only communication links delays, but also heterogeneity between vehicles' dynamics in the same string. This should be targeted ensuring the gap-regulation robustness without degrading the expected performance to keep car-following benefits (individual and string stability). In particular, when a heterogeneous string is formed, the differences between vehicles dynamics introduce disturbances in the closed loop system affecting the string stability. In [22] we presented an online Cooperative Adaptive Cruise Control (CACC) feedforward adaptation with a fractional-order feedback controller for stable heterogeneous strings of vehicles. Simulations demonstrate the advantages over conventional homogeneous structures as well as system's capability to both enhance stability and guarantee string stability regardless the vehicles distribution.

2) Design a modular architecture that permits to introduce cooperative string driving in urban environments, where interaction with vulnerable road users is highly probable. In this context, a cooperative car-following/emergency braking system with prediction-based pedestrian avoidance capabilities using vehicle-to-vehicle and vehicle-to-pedestrian communication links has been proposed in [14] and validated with RITS platforms.

3) Further extend the benefits of Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) applications on traffic flow and safety, having strict $\mathcal{L}_2$ string stability as a hard constraint, employing different calculus techniques for the control design task. A fractional-order-based control algorithm is employed to enhance the car-following and string stability performance for both ACC and CACC vehicle strings, including communication temporal delay effects has been presented in [15]. Simulation and real experiments have been conducted for validating the approach.

The aforementioned contributions have been developed in the framework of the VALET project ANR-15-CE22-0013. They have been also implemented in the vehicle platforms of RITS team, for the sake of validation and further demonstration of the final VALET system.

This scientific work can be found as well in the thesis manuscript of Carlos Flores entitled "Control Architecture for Adaptive and Cooperative Car-Following" (cf. [8]).

## 7.13. Stability analysis for controller switching in autonomous vehicles

**Participants:** Francisco Navas, Imane Mahtout, Fawzi Nashashibi.

This work investigates the Youla-Kucera (YK) parameterization to provide stable responses for autonomous vehicles when dynamics or environmental changes occur. This work explores the use of the YK parameterization in dynamics systems such as vehicles, with special emphasis on stability when some dynamic change or the traffic situation demands controller reconfiguration:

- YK parameterization provides all stabilizing controllers for a given plant. This is used in order to perform stable controller reconfiguration. Different YK-based control structures are obtained for dealing with problems such order complexity, plant disconnection or matrix inversability. Stability properties are preserved even if different structures are employed, but transient behavior between controllers changes depending on the employed YK-based structure. One of the structures presents the best transient behavior without oscillations, a lower order controller complexity and no need to disconnect the initial controller, which would be important if the system shutdown is very expensive, or the initial controller is part of a safety circuit [28]. This structure is used together with CACC applications improving CACC state-of-the-art. An hybrid behavior between two CACC controllers with different time gaps is explored by means of the YK parameterization, in order to avoid ACC degradation when communication link with preceding vehicle is lost. The proposed system uses YK parameterization and communication with a vehicle ahead (different from the preceding one) providing stable responses and, more interestingly, reducing intervehicle distances in comparison with an ACC degradation. A similar idea of hybrid behavior between CACC controller with different time gap is developed for entering/exiting vehicles in the string. In that case, YK parameterization is able to ensure stability of these merging/splitting maneuvers.
- Dual YK parameterization provides all the plants stabilized by a controller. This is employed for solving CL identification problems, or adaptive control solutions, which integrate identification and controller reconfiguration processes. YK-based CL identification uses classical OL identification algorithms, providing better results than if it is used alone. Results in a CACC-equipped vehicle prove how CL nature of the data affects a classical OL identification algorithm, and how dual YK parameterization helps to mitigate these effects. Finally, an adaptive control application is developed by using MMAC. Longitudinal dynamics of two vehicles in a CACC string are estimated within a model set, so the good CACC sytem can be chosen even if a heterogeneous string of vehicles is considered. Dynamics estimation results much more faster than other estimation processes in the literature.
- Different types of controllers and structures are used throughout Francisco Navas thesis ([10]), proving the adaptability of the YK parameterization to any type of controller. Simulation and experimental results demonstrate real implementation of stable controller reconfiguration, CL identification and adaptive control solutions dealing with dynamics changes or different traffic situations. The author thinks that YK is a suitable control framework able to ensure responses in autonomous driving.
- In [27] a design and implementation of a novel lateral control approach is proposed within Imane Mahtout thesis work. The control strategy is based on Youla-Kucera parametrization to switch gradually between controllers that are designed separately for big and small lateral errors. The presented approach studies the critical problem of initial lateral error in line following. It ensures smooth and stable transitions between controllers and provides a smooth vehicle response regardless of the lateral error. For an initial validation the work was tested in simulation, implementing a dynamic bicycle model. It has also been tested in real platforms implementing an electric Renault ZOE, with good results when activating the system at different lateral errors. Current work is focused on using YK-parametrization in estimating lateral vehicle dynamics.

## 7.14. Belief propagation inference for traffic prediction

**Participant:** Jean-Marc Lasgouttes.

This work [45], [44], in collaboration with Cyril Furtlehner (TAU, Inria), deals with real-time prediction of traffic conditions in a urban setting with incomplete data. The main focus is on finding a good way to encode available information (flow, speed, counts,...) in a Markov Random Field, and to decode it in the form of real-time traffic reconstruction and prediction. Our approach relies in particular on the Gaussian belief propagation algorithm.

This year, continuing our collaboration with PTV Sistema, we improved our techniques and obtained extensive results on large-scale datasets containing 250 to 2000 detectors. The results show very good ability to predict flow variables and a reasonably good performance on speed or occupancy variables. Some element of understanding of the observed performance are given by a careful analysis of the model, allowing to some extent to disentangle modelling bias from intrinsic noise of the traffic phenomena and its measurement process [35].

## 7.15. Large scale simulation interfacing

**Participant:** Jean-Marc Lasgouttes.

The SINETIC FUI project aims to build a complete simulation environment handling both mobility and communication. We are interested here in a so-called system-level view, focusing on simulating all the components of the system (vehicle, infrastructure, management center, etc.) and its realities (roads, traffic conditions, risk of accidents, etc.). The objective is to validate the reference scenarios that take place on a geographic area where a large number of vehicles exchange messages using the IEEE 802.11p protocol. This simulation tool is done by coupling the SUMO microscopic simulator and the ns-3 network simulator thanks to the simulation platform iTETRIS.

We have focused in this part of the project on how to reduce the execution time of large scale simulations. To this end, we designed a new simulation technique called Restricted Simulation Zone which consists on defining a set of vehicles responsible of sending the message and an area of interest around them in which the vehicles receive the packets [31].

## 7.16. Platoons Formation for autonomous vehicles redistribution

**Participants:** Mohamed Hadded, Jean-Marc Lasgouttes, Fawzi Nashashibi, Ilias Xydias.

In this paper, we consider the problem of vehicle collection assisted by a fleet manager where parked vehicles are collected and guided by fleet managers. Each platoon follows a calculated and optimized route to collect and guide the parked vehicles to their final destinations. The Platoon Route Optimization for Picking up Automated Vehicles problem, called PROPAV, consists in minimizing the collection duration, the number of platoons and the total energy required by the platoon leaders. We propose a formal definition of PROPAV as an integer linear programming problem, and then we show how to use the Non-dominated Sorting Genetic Algorithm II (NSGA-II), to deal with this multi-criteria optimization problem. Results in various configurations are presented to demonstrate the capabilities of NSGA-II to provide well-distributed Pareto-front solutions.

This work has been presented at ITSC 2018 conference [24].

## 7.17. Prediction-based handover between VLC and IEEE 802.11p for vehicular environment

**Participants:** Mohammad Abualhoul, Fawzi Nashashibi.

Despite years of development and deployment, the standardized IEEE 802.11p communication for vehicular networks can be pushed toward insatiable performance demands for wireless network data access, with a remarkable increase of both latency and channel congestion levels when subjected to scenarios with a very high vehicle density.

In specific hard safety applications such as convoys, IEEE 802.11p could seriously fail to meet the fundamental vehicular safety requirements. On the other hand, the advent of LED technologies has opened up the possibility of leveraging the more robust Visible Light Communication (VLC) technology to assist IEEE 802.11p and provide seamless connectivity in dense vehicular scenarios.

In this particular research, we proposed and validated a Prediction-based Vertical handover (PVHO) between VLC and IEEE 802.11p meant to afford seamless switching and ensure the autonomous driving safety requirements [19].

Algorithm validation and platoon system performance were evaluated using a specially implemented IEEE 802.11p-VLC module in the NS3 Network Simulator. The simulation results showed a speed-based dynamic redundancy before and after VLC interruptions with seamless switching. Moreover, the deployment of VLC for platoon intra-communication can achieve a 10-25% PDR gain in high-density vehicular scenarios, where the work was published in the IEEE International Conference on Intelligent Transportation Systems 2018.

## 7.18. Lane-Centering to Ensure the Visible Light Communication (VLC) Connectivity for a Platoon of Autonomous Vehicles

**Participants:** Mohammad Abualhoul, Fawzi Nashashibi.

VLC technology limitations were defined and supported by different solutions proposals to enhance the crucial alignment and mobility limitations. In this research [17], we proposed the incorporation of the VLC technology and a Lane-Centering (LC) technique to assure the VLC-connectivity by keeping the autonomous vehicle aligned to the lane center using vision-based lane detection in a convoy-based formation. Such combination can ensure the optical communication connectivity. This contribution by RITS-Team won the best paper award during the ICVES conference.

The system performance and evaluation showed that ss soon as the road lanes are detectable, the evaluated results showed stable behavior independently from the inter-vehicle distances and without the need for any exchanged information of the remote vehicles. Further investigations are to be carried-out in this direction.

## 7.19. Cyberphysical Constructs for Next-Gen Vehicles and Autonomic Vehicular Networks

**Participant:** Gérard Le Lann.

Behaviors of Connected Automated Vehicles (CAVs) rest on robotics capabilities (sensors, motion control laws, actuators) and wireless radio communications. Reduction of non-harmful crashes and fatalities despite higher vehicular density (safety and efficiency properties) is a fundamental objective, whatever the SAE automated driving levels considered (use cases).

Based on "hard sciences", onboard robotics capabilities designed so far are satisfactory for numerous settings, to the exception of non-line-of-sight scenarios. That is the rationale for wireless radio communications. Over the years, a growing fraction of the scientific community has been questioning the adequacy of current IEEE and ETSI standards aimed at automotive wireless communications, herein referred to as wave protocols (wireless access in vehicular environment) for convenience.

Analyses based on well-known results in various areas such as life/safety-critical systems, distributed algorithms, dependable real-time computing, ad hoc mobile networking, and cyber-physics (to name a few) come to the conclusion that wave protocols do not meet essential requirements regarding safety, efficiency, privacy or cybersecurity (SPEC). These conclusions are based on scientific demonstrations. Notably, wave protocols rest on intuitive designs (no proofs, only simulations or experimental testing) that violate well-known impossibility results in asynchronous or synchronous systems. It follows that future vehicles shall be commanded and controlled by onboard robotics supplemented with wireless communication capabilities other than wave protocols. These vehicles are referred to as Next-Gen Vehicles (NGVs) in order to avoid confusion with CAVs.

That wave solutions are far from being convincing is at the core of the recommendations issued at the EU level (the latest WG29 resolution). Moreover, the important question of how to instantiate the EU GDPR directive in future CAVs is left unanswered, despite the fact that it is possible (proofs provided) to achieve safety and privacy jointly. Preliminary results for NGVs have appeared in [34].

The work reported herein, started in 2017 along with international researchers, aims at specifying solutions to the SPEC problem, considering self-organizing and self-healing Autonomic Vehicular Networks (AVNs) of NGVs. Parallel to this, risks of privacy breaches and cyberattacks proper to wave solutions have been exposed to the public via invited interventions and presentations.

An issue not very well addressed so far is to which extent robotics and computer science supplement each other. The cyber-physical perspective is essential to formulate a coherent vision. In cyber space and in physical space, safety has to do with resource sharing. Deadlock-free and fair resource sharing in systems of concurrent processes has been a major topic in computer science for more than 50 years. Asphalt (2D systems), asphalt and air space (3D systems) are the shared resources of interest in the physical space.

As is well known, there are three classes of algorithmic solutions: detection-and-recovery, prevention, avoidance. The former class is inapplicable (one cannot "roll back an accident"). Prevention is aimed at prohibiting the emergence of hazardous (no safety) or deadlock-prone (no safety, no efficiency) conditions. Solutions are the province of distributed algorithms (computer science). Avoidance is relied on for maintaining non-hazardous conditions while making progress (also, in case some of the assumptions that underlie prevention schemes would be violated). Solutions are the province of automation control (linear/non-linear dynamics).

Prevention and avoidance schemes are needed, put in action as follows: NGVs run (cyber) distributed agreement algorithms in order to preclude the emergence of hazardous conditions, prior to executing physical motions (collision-free trajectories), which motions are made feasible thanks to prevention schemes. This is how computer science and robotics can be "married" consistently: with prevention schemes, one achieves proactive safety, and with avoidance schemes, one achieves reactive safety (both types are needed).

NGVs and AVNs are life/safety-critical cyber-physical systems. Consequently, correct solutions to the SPEC problem are based on cyber-physical constructs endowed with appropriate intrinsic properties. We have devised the cell and the cohort constructs, which rest on the obvious observation according to which only vehicles sufficiently close to each other may experience a collision. Time-bounded ultra-fast message-passing and inter-vehicular coordination can be achieved within these constructs thanks to very short-range radio and optical communications, as well as deterministic protocols (MAC protocols in particular) and distributed algorithms (dissemination, approximate agreement, and consensus). Analytical expressions of upper bounds for message-passing and inter-vehicular coordination are established for worst-case conditions, such as contention and failures, message losses in particular. We have shown that these solutions can sustain message loss frequencies an order of magnitude higher than frequencies beyond which none of the wave protocols could work.

We have defined the concept of cyberphysical levels, which are orthogonal to SAE automated driving levels. Joining a cohort longitudinally or laterally (which implies a lane change) is conditioned on a number of criteria, such as cyberphysical levels, NGV sizes, and proof of authentication (requestor's name must be a certified pseudonym).

Naming raises open problems in spontaneous mobile open systems, such as AVNs. Privacy-preserving naming is even more complex. The "longitudinal privacy-preserving naming" problem is solved with the cohort construct. The "lateral privacy-preserving naming" problem which arises with NGVs members of a cell or that circulate in adjacent cohorts has solutions based on combined optical and radio communications.

Novel deterministic time-bounded MAC protocols at the core of distributed coordination algorithms are needed to solve the open problem of safe entrances into unsignalized intersections of arbitrary topologies (any number of arterials, any number of lanes per arterial) in the presence of noisy radio channels. This problem has been solved with CSMA-CD/DCR (deterministic collision resolution) MAC protocols.

## 7.20. Functional equations

**Participant:** Guy Fayolle.

The article [13] presents functional equations (involving one or two complex variables) as an Important analytic method in stochastic modelling and in combinatorics.

## 7.21. Optimization of test case generation for ADAS via Gibbs sampling algorithms

**Participant:** Guy Fayolle.

Validating Advanced Driver Assistance Systems (ADAS) is a strategic issue, since such systems are becoming increasingly widespread in the automotive field.

But ADAS validation is a complex issue, particularly for camera based systems, because these functions may be facing a very high number of situations that can be considered as infinite. Building at a low cost level a sufficiently detailed campaign is thus very difficult. Indeed, test case generation faces the crucial question of *inherent combinatorial explosion*. An important constraint is to generate *almost all* situations in the most economical way. This task, in general, can be considered from two points of view: deterministic via binary search trees, or stochastic via Markov chain Monte Carlo (MCMC) sampling. We choose the latter probabilistic approach described below, which in our opinion seems to be the most efficient one. Typically, the problem is to produce samples of large random vectors, the components of which are possibly dependent and take a finite number of values with some given probabilities.The following flowchart is proposed.

1. In a first step, starting from the simulation graph generated by the toolboxes of MATLAB, we construct a so-called *Markov Random Field (MRF)* . When the parameters are locally dependent, this can be achieved from the user's specifications and by a systematic application of Bayes' formula.
2. Then, to cope with the combinatorial explosion, test cases are produced by implementing (and comparing) various *Gibbs samplers*, which are fruitfully employed for large systems encountered in physics. In particular, we strive to make a compromise between the convergence rate toward equilibrium, the percentage of generated duplicates and the path coverage, recalling that the speed of convergence is exponential, a classical property deduced from the general theory of Markov chains.
3. The problem of generating rare events by mixing Gibbs samplers, Large Deviation Techniques (LDT) and cross-entropy method a work in progress.

The French car manufacturer *Groupe PSA* shows a great interest in these methods and has established a contractual collaboration involving ARMINES-Mines ParisTech (Guy Fayolle as associate researcher) and Can Tho University in Vietnam (Pr. Van Ly Tran).

## 7.22. Random walks in orthants and lattice path combinatorics

**Participant:** Guy Fayolle.

In the second edition of the book [2], original methods were proposed to determine the invariant measure of random walks in the quarter plane with small jumps (size 1), the general solution being obtained via reduction to boundary value problems. Among other things, an important quantity, the so-called *group of the walk*, allows to deduce theoretical features about the nature of the solutions. In particular, when the *order* of the group is finite and the underlying algebraic curve is of genus 0 or 1, necessary and sufficient conditions have been given for the solution to be rational, algebraic or $D$-finite (i.e. solution of a linear differential equation). In this framework, number of difficult open problems related to lattice path combinatorics are currently being explored, in collaboration with A. Bostan and F. Chyzak (project-team SPECFUN, Inria-Saclay), both from theoretical and computer algebra points of view: concrete computation of the criteria, utilization of differential Galois theory, genus greater than 1 (i.e. when some jumps are of size $\geq 2$), etc. A recent topic of future research deals with the connections between simple product-form stochastic networks (so-called *Jackson networks*) and explicit solutions of functional equations for counting lattice walks.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

**Valeo Group:** a very strong partnership is under reinforcement between Valeo and Inria. Several bilateral contracts were signed to conduct joint works on Driving Assistance, some of which Valeo is funding. This joint research includes:

- The PhD thesis of Pierre de Beaucorps under the framework of Valeo project "Daring"
- A CIFRE like PhD thesis is ongoing between Valeo and Inria (Maximilian JARITZ), dealing with multisensor processing and learning techniques for free navigable road detection.
- Valeo is currently a major financing partner of the "GAT" international Chaire/JointLab in which Inria is a partner. The other partners are: UC Berkeley, Shanghai Jiao-Tong University, EPFL, IFSTTAR, MPSA (Peugeot-Citroën) and SAFRAN.
- Technology transfer is also a major collaboration topic between RITS and Valeo as well as the development of a road automated prototype.
- Finally, Inria and Valeo are partners of the PIA French project CAMPUS ( Connected Automated Mobility Platform for Urban Sustainability) including SAFRAN, Invia and Gemalto. The aim of the project is the development of autonomous vehicles and the realization of two canonical uses-cases on highways and urban like environments.

**Renault Group**: Collaboration between Renault and RITS re-started in 2016. Different research teams in Renault are now working separately with RITS on different topics.

- A CIFRE like PhD thesis is ongoing between Renault and Inria (Farouk GHALLABI) The thesis deals with the accurate localization of an autonomous vehicle on a highway using mainly on-board low-cost perception sensors.
- Another CIFRE PhD thesis is ongoing since November 2017 (Imane MAHTOUT).

**AKKA Technologies**: Collaboration with AKKA since 2012 (for the Link & Go prototype).

- Inria and AKKA Technologies are partners in the VALET projects (ANR projects).
- A CIFRE PhD thesis (Luis ROLDAO) dealing with 3D-environment modeling for autonomous vehicles begun in October 2017.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. ANR*

*9.1.1.1. VALET*

Title: Redistribution automatique d'une flotte de véhicules en partage et valet de parking

Instrument: ANR

Duration: January 2016 - December 2018

Coordinator: Fawzi Nashashibi

Partners: Inria, Ecole Centrale de Nantes (IRCCyN), AKKA Technologies

Inria contact: Fawzi Nashashibi

Abstract: The VALET project proposes a novel approach for solving car-sharing vehicles redistribution problem using vehicle platoons guided by professional drivers. An optimal routing algorithm is in charge of defining platoons drivers' routes to the parking areas where the followers are parked in a complete automated mode. The main idea of VALET is to retrieve vehicles parked randomly on the urban parking network by users. These parking spaces may be in electric charging stations, parking for car sharing vehicles or in regular parking places. Once the vehicles are collected and guided in a platooning mode, the objective is then to guide them to their allocated parking area or to their respective parking lots. Then each vehicle is assigned a parking place into which it has to park in an automated mode.

*9.1.1.2. Hianic*

Title: navigation autonome dans les foules inspirée par les humains (Human Inspired Autonomous Navigation In Crowds)
Instrument: ANR
Duration: January 2018 - December 2020
Coordinator: Anne Spalanzani (Inria Rhône-Alpes, Chroma research team)
Partners: Inria Rhône-Alpes, Inria Paris, LIG Laboratoire d'Informatique de Grenoble, LS2N - ECN Laboratoire des Sciences du Numérique de Nantes
Inria contact: Fawzi Nashashibi
Abstract: The HIANIC project will try to address some problems that will arise when these cars are mixed with pedestrians. The HIANIC project will develop new technologies in term of autonomous navigation in dense and human populated traffic. It will explore the complex problem of navigating autonomously in shared-space environments, where pedestrians and cars share the same environment.

Such a system will contribute both to urban safety and intelligent mobility in "shared spaces". Negotiation will help to avoid frozen situations increasing the vehicle's reactivity and optimizing the navigable space. Negotiation, Human-Aware Navigation and Communication will contribute to a better public acceptance of such autonomous systems and facilitate their penetration in the transportation landscape.

### *9.1.2. FUI*

#### *9.1.2.1. Sinetic*

Title: Système Intégré Numérique pour les Transports Intelligents Coopératifs
Instrument: FUI
Duration: December 2014 - January 2018
Coordinator: Thomas Nguyen (Oktal)
Partners: Oktal, ALL4TEC, CIVITEC, Dynalogic, Inria, EURECOM, Renault, Armines, IFSTTAR, VEDECOM
Inria contact: Jean-Marc Lasgouttes
Abstract: The purpose of the project SINETIC is to create a complete simulation environment for designing cooperative intelligent transport systems with two levels of granularity: the system level, integrating all the components of the system (vehicles, infrastructure management centers, etc.) and its realities (terrain, traffic, etc.) and the component-level, modeling the characteristics and behavior of the individual components (vehicles, sensors, communications and positioning systems, etc.) on limited geographical areas, but described in detail.

#### *9.1.2.2. PAC V2X*

Title: Perception augmentée par coopération véhicule avec l'infrastructure routière
Instrument: FUI
Duration: September 2016 - August 2019
Coordinator: SIGNATURE Group (SVMS)
Partners: DigiMobee, LOGIROAD, MABEN PRODUCTS, SANEF, SVMS, VICI, Inria, VEDECOM
Inria contact: Raoul de Charette
Abstract: The objective of the project is to integrate two technologies currently being deployed in order to significantly increase the time for an automated vehicle to evolve autonomously on European road networks. It is the integration of technologies for the detection of fixed and mobile objects such as radars, lidars, cameras ... etc. And local telecommunication technologies for the development of ad hoc local networks as used in cooperative systems.

### *9.1.3. Competitivity Clusters*

RITS team is a very active partner in the competitivity clusters, especially MOV'EO and System@tic. We are involved in several technical committees like the DAS SUR of MOV'EO for example.

RITS is also the main Inria contributor in the VEDECOM institute (IEED). VEDECOM financed the PhD theses of Mr. Fernando Garrido and Mr. Zayed Alsayed.

## 9.2. European Initiatives

### *9.2.1. FP7 & H2020 Projects*

#### *9.2.1.1. AUTOCITS*

Title: AUTOCITS Regulation Study for Interoperability in the Adoption of Autonomous Driving in European Urban Nodes

Program: CEF- TRANSPORT Atlantic corridor

Duration: November 2016 - March 2019

Coordinator: Indra Sistemas S.A. (Spain)

Partners: Indra Sistemas S.A. (Spain); Universidad Politécnica de Madrid (UPM), Spain; Dirección General de Tráfico (DGT), Spain; Inria (France); Instituto Pedro Nunes (IPN), Portugal; Autoridade Nacional de Segurança Rodoviária (ANSR), Portugal; Universidade de Coimbra (UC), Portugal.

Inria contact: Fawzi Nashashibi, Mohammad Abualhoul

Abstract: The aim of the Study is to contribute to the deployment of C-ITS in Europe by enhancing interoperability for autonomous vehicles as well as to boost the role of C-ITS as catalyst for the implementation of autonomous driving. Pilots will be implemented in 3 major Core Urban nodes (Paris, Madrid, Lisbon) located along the Core network Atlantic Corridor in 3 different Member States. The Action consists of Analysis and design, Pilots deployment and assessment, Dissemination and communication as well as Project Management and Coordination.

### *9.2.2. Collaborations with Major European Organizations*

RITS is member of the **euRobotics AISBL** (Association Internationale Sans But Lucratif) and the Leader of "People transport" Topic. This makes from Inria one of the rare French robotics representatives at the European level. See also: http://www.eu-robotics.net/

RITS is a full partner of **VRA – Vehicle and Road Automation**, a support action funded by the European Union to create a collaboration network of experts and stakeholders working on deployment of automated vehicles and its related infrastructure. VRA project is considered as the cooperation interface between EC funded projects, international relations and national activities on the topic of vehicle and road automation. It is financed by the European Commission DG CONNECT and coordinated by ERTICO – ITS Europe. See also: http://vra-net.eu/

## 9.3. International Initiatives

### *9.3.1. Inria International Partners*

#### *9.3.1.1. Informal International Partners*

RITS has signed 3 MoU with the following international laboratories:

- Vehicle Dynamics and Control Laboratory, Seoul National University (SNU), S. Korea: international cooperation agreement for Graduate-Level Academic and Research Collaboration
- MICA Lab, Hanoi University of Science and Technology, Vietnam: cooperation agreement for research collaboration and PhD students co-supervision
- Integrated Industrial Design Lab (INDEL) of the Department of Product and Systems Design Engineering, University of the Aegean, Greece: international cooperation agreement for Graduate-Level Academic and Research Collaboration

### 9.3.2. Participation in International Programs

Samuel de Champlain Québec-France collaboration program: "Vision par ordinateur en conditions difficiles", cooperation between Raoul de Charette and Jean-François Lalonde from Laval University.

## 9.4. International Research Visitors

### 9.4.1. Visits of International Scientists

Jean-François Lalonde from Laval University in October 2018 within the framework of Samuel de Champlain Québec-France collaboration program.

*9.4.1.1. Internships*

Shirsendu Halder, June-December 2018.

Nabila Arib, April-September 2018

### 9.4.2. Visits to International Teams

*9.4.2.1. Research Stays Abroad*

Maximilian Jaritz was at UC San Diego, visiting SU Lab directed by Hao Su, from October 1st 2018 to February 15th 2019.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### 10.1.1. Scientific Events Organisation

*10.1.1.1. General Chair, Scientific Chair*

Mohammad Abualhoul and Fawzi Nashashibi: chairs and co-organizers of the 1st CA2V workshop at the IEEE VTC'18 conference, Porto (Portugal), June 2018.

Fawzi Nashashibi and Mohammad Abualhoul: chairs and co-organizers of the CA2V workshop at the IEEE ICVES'18 conference, Madrid (Spain), September 2018.

Fawzi Nashashibi: chair of the AUTOCITS international workshop ob connected and autonomous vehicles, December 13, Paris, France.

*10.1.1.2. Member of the Organizing Committees*

Fawzi Nashashibi was member of the organizing committee of the IEEE/RSJ IROS 2018 Autonomous Driving Events including: PPNIV Workshop, Round table and Autonomous vehicle on-site demonstrations.

Fawzi Nashashibi was member of the Steering Committee of the IEEE Global Communications Conference, 9-13 December 2018, Abu Dhabi, UAE.

### 10.1.2. Scientific Events Selection

Jean-Marc Lasgouttes: co-organizer of the Symposium "Large Random Networks and Constrained Random Walks" in honor of the $75^{\text{th}}$ birthday of Guy Fayolle, Dijon (France), August 27–28.

*10.1.2.1. Chair of Conference Program Committees*

Raoul de Charette: program chair of Workshop 'Real-World Challenges for Robotic Vision' CVPR 2018 (CVPRW).

*10.1.2.2. Member of the Conference Program Committees*

Raoul de Charette: program committee member of 'Conférence Française de Photogrammétrie et de Télédétection' (CFPT).

Fawzi Nashashibi : co-organizer of the breakout session *Artificial Intelligence and Deep Machine Learning Tools and Algorithms for Automated Vehicles: The State of the Art and Practice* at the Automated Vehicles Symposium, San Francisco (USA), 9-12 July, 2018.

*10.1.2.3. Reviewer*

Mohammad Abualhoul: *IEEE ITS 2018*, *IEEE IV 2018*, *IEEE ICVES 2018*, *TRA 2018*.

Raoul de Charette: *CVPR 2018*, *ACCV 2018*, *IEEE ICRA 2018*, *CVPR Workshop 2018*, *IEEE IV 2018*.

Jean-Marc Lasgouttes: *IEEE ITSC 2018*.

Fawzi Nashashibi : *IEEE ICRA 2018*, *IEEE IROS 2018*, *IEEE IV 2018*, *IEEE ITSC 2018*, *IEEE ICARCV 2018*, *IEEE VTC 2018*, *IEEE ICVES 2018*

Anne Verroust-Blondet: *IEEE IV 2018*, *IEEE ITSC 2018*, *IEEE ICARCV 2018*.

### *10.1.3. Journal*

*10.1.3.1. Member of the Editorial Boards*

Guy Fayolle: associate editor of the journal *Markov Processes and Related Fields*.

Fawzi Nashashibi: associate editor of the journal *IEEE Transactions on Intelligent Vehicles*, associate editor of the journal *IEEE Transactions on Intelligent Transportation Systems*.

Anne Verroust-Blondet: associate editor of the journal *The Visual Computer*.

*10.1.3.2. Reviewer - Reviewing Activities*

Raoul de Charette: *IEEE Transactions on Intelligent Transportation Systems*

Guy Fayolle: *AAP, MPRF, PTRF, QUESTA, European Journal of Combinatorics, JSP, Physica A, Springer Science*.

Jean-Marc Lasgouttes: *IEEE Transactions on Intelligent Transportation Systems*

Fawzi Nashashibi: *IEEE Transactions on Intelligent Transportation Systems*, *IEEE Transactions on Intelligent Vehicles*, *Transportation Research Part C*.

Anne Verroust-Blondet: *IEEE Transactions on Intelligent Vehicles*

### *10.1.4. Invited Talks*

Mohammad Abualhoul: keynote speaker in ICVES-CA2V-2018 Worshop in Madrid, June 2018.

Raoul de Charette: public presentation at BNF Paris on Artificial Intelligence (2018-10-08).

Guy Fayolle: guest speaker at the conference *Grands réseaux aléatoires et marches contraintes*, held at the University of Bourgogne, Dijon, 27-28th August 2018, http://www.lmpt.univ-tours.fr/ConferenceGuyFayolle75/.

Guy Fayolle: four long talks on topics related to the book [2] at the working group GTMA (http://divizio.joomla.com/seminaires-et-gdt/11-groupe-de-travail-autour-des-marches-dans-le-quart-de-plan).

Gérard Le Lann: "Conduite automatisée, cyber-sécurité et enjeux de société", Espace des Sciences, Rennes, April 2018.

Gérard Le Lann: "Future Directions in Autonomic Vehicular Networks", IEEE & IFIP Med-Hoc-Net Workshop, Capri, Italy, June 2018.

Gérard Le Lann: "Future Directions in Connected Autonomous Vehicles and Autonomic Vehicular Networks", ATOS Worldwide Annual TechForum, Seclin, July 2018.

Gérard Le Lann: "Future Directions in Connected Autonomous Vehicles and Autonomic Vehicular Networks", CNIL, Paris, September 2018.

Gérard Le Lann: "Safety, Privacy and Cybersecurity in Future Vehicular Networks", AutoCITS Workshop, Inria Paris, December 2018.

Fawzi Nashashibi: keynote speaker on *Scientific and technical challenges of autonomous navigation for connected autonomous vehicles* in ICVES-CA2V-2018 Worshop in Madrid, June 2018.

Fawzi Nashashibi: keynote speaker on *Autonomous vehicles in smart cities: the next challenges* in PAMS international conference, Pangyo (S. Korea), 15-16 November 2018.

Fawzi Nashashibi: keynote and Round table speaker on *Scientific and technical challenges of autonomous navigation for connected autonomous vehicles* at the International Conference on Mobility Challenges, Gif-sur-Yvette (France), 6-7 December 2018.

Fawzi Nashashibi: keynote on *Evolution and challenges of mobility with the increased introduction of artificial intelligence*, at the TOTAL AI FOR LEADERS PROGRAM, Paris (France), March 19, 2018.

Fawzi Nashashibi: keynote on the *Challenges et opportunities of autonomous mobility in smart cities*, in the working day on Intelligent Mobility, organized by TECNALIA, Madrid (Spain), February 2018.

### *10.1.5. Scientific Expertise*

Guy Fayolle is scientific advisor and associate researcher at the *Robotics Laboratory of Mines ParisTech*. He is also collaborating member of the research-team SPECFUN at Inria-Saclay.

Jean-Marc Lasgouttes is member of the *Conseil Académique* of Université Paris-Saclay.

Gérard Le Lann: contributions to *Report on Safer Roads with Automated Vehicles ?*, International Transportation Forum, OECD, April 2018, 45 p. [0]

Fawzi Nashashibi is an associate researcher at the *Robotics Laboratory, Mines ParisTech*. He is an evaluator/reviewer of European H2020 projects.

### *10.1.6. Research Administration*

Jean-Marc Lasgouttes is a member of the *Comité Technique Inria*.

Guy Fayolle is a member of the working group IFIP WG 7.3.

Fawzi Nashashibi is a member of the international Automated Highway Board Committee of the TRB (AHB30). He is a member of the Board of Governors of the VEDECOM Institute representing Inria and of the Board of Governors of MOV'EO Competitiveness cluster representing Inria.

Anne Verroust-Blondet is the scientific correspondent of the European affairs and of the International Partnerships for Inria Paris, member of the COST-GTRI committee at Inria (Committee in charge of the evaluation of international projects) and mamber of the "emplois scientifiques" committee of Inria Paris.

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Licence: Fawzi Nashashibi, "Programmation avancée", 84h, L1, Université Paris-8 Saint-Denis, France.

Master: Carlos Flores and Anne Verroust-Blondet, "Le véhicule autonome. Présentation des recherches de l'équipe-projet RITS",1.5 h, 2nd year, Ecole des Ponts ParisTech, France, September 2018.

Master: Jean-Marc Lasgouttes, "Analyse de données", 54h, second year of Magistère de Finance (M1), University Paris 1 Panthéon Sorbonne, France.

Master: Jean-Marc Lasgouttes, "Analyse de données", 52.5h, Master 1 SIC APP, University Paris 1 Panthéon Sorbonne, France.

[0] https://www.itf-oecd.org/sites/default/files/docs/safer-roads-automated-vehicles.pdf

Master: Fawzi Nashashibi, “Image synthesis and 3D Infographics”, 12h, M2, INT Télécom Sud-Paris.

Master: Fawzi Nashashibi, “Obstacle detection and Multisensor Fusion”, 4h, M2, INSA de Rouen.

Master: Fawzi Nashashibi, “Perception and Image processing for Mobile Autonomous Systems”, 12h, M2, University of Evry.

Master: Renaud Poncelet, “Mécatronique fondamentale”, 64h, 4rd year, Pôle Universitaire Léonard de Vinci, Paris La Défense, France.

Doctorat: Jean-Marc Lasgouttes, “Analyse de données fonctionnelles”, 31.5h, Mastère Spécialisé “Expert en sciences des données”, INSA-Rouen, France

### *10.2.2. Supervision*

PhD: Zayed Alsayed,"Characterizing the Robustness and Enhancing the Accuracy of SLAM-based Localization Systems for Autonomous Driving", Télécom ParisTech, June 2018, supervisor: Anne Verroust-Blondet, co-supervisors: Fawzi Nashashibi, Guillaume Bresson.

PhD: Carlos Flores, “Control Architecture for Adaptive and Cooperative Car-Following”, Mines ParisTech, PSL Research University, December 2018, supervisor: Fawzi Nashashibi, co-supervisor: Vicente Milanés.

PhD: Fernando Garrido, “Two-staged local trajectory planning based on optimal pre-planned curves interpolation for human-like driving in urban areas” , Mines ParisTech, PSL Research University, December 2018, supervisor: Fawzi Nashashibi, co-supervisors: Vicente Milanés, Joshué Pérez.

PhD: Francisco Navas, “Stability analysis for controller switching in autonomous vehicles”, Mines ParisTech, , PSL Research University, November 2018, supervisor: Fawzi Nashashibi, co-supervisor: Vicente Milanés.

PhD: Dinh-Van Nguyen, "Wireless Sensors Networks for Indoor Mapping and Accurate Localization for Low Speed Navigation in Smart Cities", Mines ParisTech, , PSL Research University, December 2018, supervisor: Fawzi Nashashibi, co-supervisor: Eric Castelli.

PhD in progress: Pierre de Beaucorps, “Autonomous vehicle: behavior prediction and Interaction with road users”, UPMC Sorbonne University, January 2016, supervisor: Anne Verroust-Blondet, co-supervisor: Fawzi Nashashibi.

PhD in progress: Farouk Ghallabi, “Environment modeling and simultaneous localization of a mobile vehicle on motorways: a multi-sensor approach”, Mines ParisTech, PSL Research University, October 2016, supervisor: Fawzi Nashashibi.

PhD in progress: Maximilian Jaritz, “Perception multi-capteur pour la conduite autonome grâce à l’apprentissage profond”, Mines ParisTech, PSL Research University, January 2017, supervisor: Fawzi Nashashibi, co-supervisor: Raoul de Charette.

PhD in progress: Danut-Ovidiu Pop, "Deep learning techniques for intelligent vehicles", INSA Rouen, May 2016, supervisor: Abdelaziz Bensrhair, co-supervisor: Fawzi Nashashibi.

PhD in progress: Imane Matout, “Estimation de l’intention des véhicules pour la prise de décision et le contrôle sans faille en navigation autonome”, Mines ParisTech, PSL Research University, December 2017, supervisor: Fawzi Nashashibi, co-supervisor: Vicente Milanés.

PhD in progress: Kaouther Messaoud, “Détermination des manoeuvres et des intentions des véhicules avoisinant un véhicule autonome”, UPMC Sorbonne University, October 2017, supervisor: Anne Verroust-Blondet, co-supervisors: Fawzi Nashashibi, Itheri Yahiaoui.

PhD in progress: Luis Roldao, “Modélisation 3D de l’environnement et de la manoeuvrabilité d’un véhicule”, UPMC Sorbonne University, October 2017, supervisor: Anne Verroust-Blondet, co-supervisor: Raoul de Charette.

Starting PhD: Renaud Poncelet, “Navigation autonome en présence d’obstacles fortement dynamiques au mouvement incertain”, UPMC Sorbonne University, September 2018, supervisor: Anne Verroust-Blondet, co-supervisor: Fawzi Nashashibi.

### 10.2.3. Juries

Fawzi Nashashibi was a jury member of the PhD thesis of Mr. Laurent Delobel - *Agrégation d'Information Pour la Localisation d'un Robot Mobile sur une Carte Imparfaite* -, Université Clermont-Auvergne, Clermont-Ferrand, 16 February 2018.

Fawzi Nashashibi was a reviewer of the PhD thesis of Mr. Yrvann Emzivat *Safety System Architecture for the Design of Dependable and Adaptable Autonomous Vehicles* -, Université Bretagne Loire, Nantes (France), 30 May 2018.

Fawzi Nashashibi was a jury member of the PhD thesis of Mr. Aymeric Dujardin - *Détection d'obstacles par stéréovision en environnement non structuré* -, Normandie Université, Saint-Etienne-du-Rouvray (France), 03 July 2018.

Fawzi Nashashibi participated, as examiner and president, to the HdR committee of Oyunchimeg Shagdar *Optimizing Wireless Communications in Dense Mobile Environments*, Université de Versailles Saint-Quentin-en-Yvelines, 6 September 2018.

Anne Verroust-Blondet was a reviewer of the PhD thesis of Sarah Bertrand - *Analyse d'images pour l'identification multi-organes d'espèces végétales*, Université Lyon 2, 10 December 2018.

Anne Verroust-Blondet was a jury member of the PhD thesis of Dai-Duong Nguyen - *A vision system based real-time SLAM application*, Université Paris-Saclay, 7 December 2018.

## 10.3. Popularization

### 10.3.1. Internal or external Inria responsibilities

Fawzi Nashashibi was the coordinator of the Inria Livre Blanc "Véhicules autonomes et connectés, les défis actuels et les voies de recherche" [0].

### 10.3.2. Articles and contents

A few press interviews and popular science was done in 2018.

- In books/journals/radios for the general public:

  Raoul de Charette was interviewed by A. Devillard for:

  - *Pour la voiture autonome, l'algorithme d'apprentissage qui ne s'entraîne pas*[0], Science et Avenir, 2018-03-18;

  - *Les failles de la voiture autonome*[0], Sciences et Avenir, 2018-04-26;

  Raoul de Charette and Fawzi Nashashibi were interviewed by M. Moragues for *[Infographie] L'intelligence artificielle n'est pas un as du volant*[0], L'Usine Nouvelle, 2018-09-23;

  Raoul de Charette was interviewed for Sciences et Avenir, *La voiture autonome en quête de la "vision" parfaite*, 2018-09-27;

  Raoul de Charette and Fawzi Nashashibi were interviewed for Challenges, *L'intelligence artificielle bute sur la sécurité routière*, 2018-10-04;

  Raoul de Charette was interviewed for Radio France Internationale, 2018-03-30;

  Fawzi Nashashibi was interviewed for La Tribune, *Ce que l'IA change dans l'auto : l'ère de la voiture autonome*[0], 2018-05-02;

[0] https://www.inria.fr/institut/strategie/vehicules-autonomes-et-connectes
[0] https://www.sciencesetavenir.fr/high-tech/pour-la-voiture-autonome-un-algorithme-d-apprentissage-qui-ne-s-entraine-pas_122033
[0] https://www.sciencesetavenir.fr/high-tech/transports/les-failles-de-la-voiture-autonome_123502
[0] https://www.usinenouvelle.com/editorial/infographie-l-intelligence-artificielle-n-est-pas-un-as-du-volant.N742309
[0] https://www.latribune.fr/entreprises-finance/industrie/automobile/ce-que-l-ia-change-dans-l-auto-l-ere-de-la-voiture-autonome-776767.html

Fawzi Nashashibi was interviewed by Dominique Filippone for *Voitures autonomes : Pas avant 2040 en France selon l'Inria*[0], Le Monde Informatique, 2018-05-15;

Fawzi Nashashibi was interviewed by Loup Besmond de Senneville for *Les recommandations des chercheurs pour l'éthique des voitures autonomes*[0], La Croix, 2018-05-18;

Fawzi Nashashibi was interviewed by Frank Niedercorn for *Fawzi Nashashibi : "Il y a encore bien des obstacles pour le véhicule autonome"*[0], Les Echos, 2018-06-29;

Fawzi Nashashibi was interviewed for L'Usine Nouvelle, no 3577, *Automobile le graal de l'autonomie*, 2018-09-20;

Fawzi Nashashibi was interviewed by Laurène Le Fourner for Radio Village Innovation, July 2018.

Fawzi Nashashibi was interviewed for France Info, 2018-09-07.

- For online publications: (Interstices, Images des Maths, Binaire, Wikipedia), and more widely blog articles:

  Raoul de Charette was interviewed by K. Haske for MyScienceWork *Where are they now ? Raoul de Charette*[0], 2018-09-04, and by J. Jongwane for Interstices *Les algorithmes de vision pour les véhicules autonomes*[0] [36];

  Fawzi Nashashibi was interviewed by Mrs. Barbara Vignaux to prepare an exhibition on the autonomous car for the *Science Actualités* area of the City of Science and Industry (*Cité des Sciences et des Industries*), Paris. It was followed by the development of an article entitled *"Voiture autonome: révolution en route ?"* (Autonomous Car: revolution on the way?);

  In January 2018, Gérard Le Lann has been invited to contribute to Blog Binaire Le Monde on: *Sécurité routière et cybersécurité*[0];

  In May 18th, an on-line article titled *Inria dévoile son Livre blanc sur les Véhicules Autonomes et Connectés* and relaying the contents of Inria white paper on connected autonomous vehicles was published at http://www.auto-innovations.com/communique/1374.html.

### *10.3.3. Interventions*

- Raoul de Charette and Renaud Poncelet made a demonstration at Inria RII meeting ("Mobilité & environnements", 2018-11-20, Station F, Paris);
- RITS was involved in welcoming schoolchildren in Paris center on December 21st.

# 11. Bibliography

## Major publications by the team in recent years

[1] Z. ALSAYED, G. BRESSON, A. VERROUST-BLONDET, F. NASHASHIBI.*2D SLAM Correction Prediction in Large Scale Urban Environments*, in "ICRA 2018 - International Conference on Robotics and Automation 2018", Brisbane, Australia, May 2018, https://hal.inria.fr/hal-01829091

[0] https://www.lemondeinformatique.fr/actualites/lire-voitures-autonomes-pas-avant-2040-en-france-selon-l-inria-71733.html
[0] https://www.la-croix.com/Sciences-et-ethique/Ethique/recommandations-chercheurs-lethique-voitures-autonomes-2018-05-18-1200940092
[0] https://www.lesechos.fr/29/06/2018/LesEchos/22727-506-ECH_fawzi-nashashibi—-il-y-a-encore-bien-des-obstacles-pour-le-vehicule-autonome-.htm
[0] https://www.mysciencework.com/omniscience/now-raoul-charette
[0] https://interstices.info/les-algorithmes-de-vision-pour-les-vehicules-autonomes/
[0] http://binaire.blog.lemonde.fr/2018/01/19/securite-routiere-et-cybersecurite/

[2] G. FAYOLLE, R. IASNOGORODSKI, V. A. MALYSHEV. , S. ASMUSSEN, P. W. GLYNN, Y. LE JAN (editors)*Random Walks in the Quarter Plane: Algebraic Methods, Boundary Value Problems, Applications to Queueing Systems and Analytic Combinatorics*, Probability Theory and Stochastic Modelling, Springer International Publishing, February 2017, vol. 40, 255, The first edition was published in 1999

[3] C. FLORES, P. MERDRIGNAC, R. DE CHARETTE, F. NAVAS, V. MILANÉS, F. NASHASHIBI.*A Cooperative Car-Following/Emergency Braking System With Prediction-Based Pedestrian Avoidance Capabilities*, in "IEEE Transactions on Intelligent Transportation Systems", June 2018, p. 1 - 10 [*DOI :* 10.1109/TITS.2018.2841644], https://hal.inria.fr/hal-01835121

[4] C. FLORES, V. MILANÉS.*Fractional-Order-Based ACC/CACC Algorithm for Improving String Stability*, in "Transportation research. Part C, Emerging technologies", October 2018, https://hal.inria.fr/hal-01896558

[5] D. GONZALEZ BAUTISTA, J. PÉREZ, V. MILANÉS, F. NASHASHIBI.*A Review of Motion Planning Techniques for Automated Vehicles*, in "IEEE Transactions on Intelligent Transportation Systems", April 2016 [*DOI :* 10.1109/TITS.2015.2498841], https://hal.inria.fr/hal-01397924

[6] M. JARITZ, R. DE CHARETTE, M. TOROMANOFF, E. PEROT, F. NASHASHIBI.*End-to-End Race Driving with Deep Reinforcement Learning*, in "ICRA 2018 - IEEE International Conference on Robotics and Automation", Brisbane, Australia, May 2018, https://arxiv.org/abs/1807.02371 , https://hal.inria.fr/hal-01848067

# Publications of the year

## Doctoral Dissertations and Habilitation Theses

[7] Z. ALSAYED.*Characterizing the Robustness and Enhancing the Accuracy of SLAM-based Localization Systems for Autonomous Driving*, Télécom ParisTech, June 2018

[8] C. FLORES.*Control Architecture for Adaptive and Cooperative Car-Following*, PSL Research University, December 2018

[9] F. GARRIDO.*Two-staged local trajectory planning based on optimal pre-planned curves interpolation for human-like driving in urban areas*, PSL Research University, December 2018

[10] F. NAVAS.*Stability analysis for controller switching in autonomous vehicles*, PSL Research University, December 2018

[11] D. V. NGUYEN.*Wireless Sensors Networks for Indoor Mapping and Accurate Localization for Low Speed Navigation in Smart Cities*, PSL Research University, November 2018

## Articles in International Peer-Reviewed Journal

[12] N. BOULILA, M. HADDED, A. LAOUITI, L. AZOUZ SAIDANE.*Hybrid centralised MAC protocol for reliable broadcast of safety messages in vehicular ad hoc network*, in "International Journal of Space-Based and Situated Computing", January 2018, vol. 8, n$^{o}$ 3, p. 169 - 178 [*DOI :* 10.1504/IJSSC.2018.097294], https://hal.archives-ouvertes.fr/hal-01994662

[13] G. FAYOLLE.*Functional equations as an important analytic method in stochastic modelling and in combinatorics*, in "Markov Processes And Related Fields", December 2018, vol. 24, n$^{o}$ 5, https://arxiv.org/abs/1712.02271 , https://hal.inria.fr/hal-01657154

[14] C. FLORES, P. MERDRIGNAC, R. DE CHARETTE, F. NAVAS, V. MILANÉS, F. NASHASHIBI.*A Cooperative Car-Following/Emergency Braking System With Prediction-Based Pedestrian Avoidance Capabilities*, in "IEEE Transactions on Intelligent Transportation Systems", June 2018, p. 1 - 10 [*DOI :* 10.1109/TITS.2018.2841644], https://hal.inria.fr/hal-01835121

[15] C. FLORES, V. MILANÉS.*Fractional-Order-Based ACC/CACC Algorithm for Improving String Stability*, in "Transportation research. Part C, Emerging technologies", October 2018, https://hal.inria.fr/hal-01896558

[16] K. JO, S. CHO, C. KIM, P. RESENDE, B. BRADAI, F. NASHASHIBI, M. SUNWOO.*Cloud Update of Tiled Evidential Occupancy Grid Maps for the Multi-Vehicle Mapping*, in "Sensors", December 2018, vol. 18, n° 12, 4119 [*DOI :* 10.3390/s18124119], https://hal.inria.fr/hal-01968617

### Invited Conferences

[17] *Best Paper*
M. ABUALHOUL, E. TALAVERA MUNOZ, F. NASHASHIBI.*The Use of Lane-Centering to Ensure the Visible Light Communication Connectivity for a Platoon of Autonomous Vehicles*, in "ICVES'2018 - 20th IEEE International Conference on Vehicular Electronics and Safety", Madrid, Spain, September 2018, https://hal.inria.fr/hal-01888549.

[18] R. CASTIÑEIRA, J. E. NARANJO, M. GIL, F. JIMENEZ, P. SERRA, A. VALEJO, A. ASVADI, C. PREMEBIDA, M. ABUALHOUL, F. NASHASHIBI.*AUTOCITS – Regulation study for interoperability in the adoption of autonomous driving in European urban nodes*, in "TRA 2018 - Transport Research Arena", Vienna, Austria, April 2018, https://hal.inria.fr/hal-01898256

### International Conferences with Proceedings

[19] M. ABUALHOUL, M. AL-BADO, O. SHAGDAR, F. NASHASHIBI.*A Proposal for VLC-Assisting IEEE802.11p Communication for Vehicular Environment Using a Prediction-based Handover*, in "ITSC 2018 – 21st IEEE International Conference on Intelligent Transportation Systems", Maui, Hawaii, U.S. Outlying Islands, November 2018, https://hal.inria.fr/hal-01888576

[20] Z. ALSAYED, G. BRESSON, A. VERROUST-BLONDET, F. NASHASHIBI.*2D SLAM Correction Prediction in Large Scale Urban Environments*, in "ICRA 2018 - International Conference on Robotics and Automation 2018", Brisbane, Australia, May 2018, https://hal.inria.fr/hal-01829091

[21] N. BOULILA, M. HADDED, A. LAOUITI, L. AZOUZ SAIDANE.*QCH-MAC: A Qos-aware Centralized Hybrid MAC protocol for Vehicular Ad Hoc NETworks*, in "AINA 2018: 32nd International Conference on Advanced Information Networking and Applications", Krakow, Poland, IEEE Computer Society, May 2018, p. 55 - 62 [*DOI :* 10.1109/AINA.2018.00021], https://hal.archives-ouvertes.fr/hal-01994055

[22] C. FLORES, V. MILANÉS, F. NASHASHIBI.*Online Feedforward/Feedback Structure Adaptation for Heterogeneous CACC Strings*, in "American Control Conference (ACC) 2018", Milwaukee, United States, American Control Conference (ACC) 2018, June 2018, https://hal.inria.fr/hal-01835132

[23] F. GHALLABI, F. NASHASHIBI, G. EL-HAJ-SHHADE, M.-A. MITTET.*LIDAR-Based Lane Marking Detection For Vehicle Positioning in an HD Map*, in "2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC)", Maui, Hawaii,, United States, 2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC), November 2018, https://hal.archives-ouvertes.fr/hal-01891764

[24] M. HADDED, J.-M. LASGOUTTES, F. NASHASHIBI, I. XYDIAS.*Platoon Route Optimization for Picking up Automated Vehicles in an Urban Network*, in "ITSC 2018 - 21st IEEE International Conference on Intelligent Transportation Systems", Maui, United States, 2018 IEEE 21th International Conference on Intelligent Transportation Systems (ITSC), November 2018, https://hal.inria.fr/hal-01880388

[25] M. JARITZ, R. DE CHARETTE, M. TOROMANOFF, E. PEROT, F. NASHASHIBI.*End-to-End Race Driving with Deep Reinforcement Learning*, in "ICRA 2018 - IEEE International Conference on Robotics and Automation", Brisbane, Australia, May 2018, https://arxiv.org/abs/1807.02371 , https://hal.inria.fr/hal-01848067

[26] M. JARITZ, R. DE CHARETTE, E. WIRBEL, X. PERROTTON, F. NASHASHIBI.*Sparse and Dense Data with CNNs: Depth Completion and Semantic Segmentation*, in "3DV 2018 - 6th international conference on 3D Vision", Verona, Italy, 3DV 2018 - 6th international conference on 3D Vision, September 2018, https://arxiv.org/abs/1808.00769 , https://hal.archives-ouvertes.fr/hal-01858241

[27] I. MAHTOUT, F. NAVAS, D. GONZALEZ BAUTISTA, V. MILANÉS, F. NASHASHIBI.*Youla-Kucera Based Lateral Controller for Autonomous Vehicle*, in "ITSC 2018 - 21st IEEE International Conference on Intelligent Transportation Systems", Maui, Hawaii, U.S. Outlying Islands, November 2018, https://hal.inria.fr/hal-01906268

[28] F. NAVAS, I. MAHTOUT, V. MILANÉS, F. NASHASHIBI.*Youla-Kucera control structures for switching*, in "CCTA 2018 - 2nd IEEE Conference on Control Technology and Applications", Copenhagen, Denmark, August 2018, https://hal.inria.fr/hal-01801224

[29] D. V. NGUYEN, R. DE CHARETTE, T.-K. DAO, E. CASTELLI, F. NASHASHIBI.*WiFi Fingerprinting Localization for Intelligent Vehicles in Car Park*, in "IPIN 2018 : Ninth International Conference on Indoor Positioning and Indoor Navigation", Nantes, France, September 2018, https://hal.inria.fr/hal-01851504

[30] B. RAVI KIRAN, L. ROLDAO, B. IRASTORZA, R. VERASTEGUI, S. SÜSS, S. YOGAMANI, V. TALPAERT, A. LEPOUTRE, G. TREHARD.*Real-time Dynamic Object Detection for Autonomous Driving using Prior 3D-Maps*, in "First International Workshop On Autonomous Navigation in Unconstrained Environments - In Conjunction with ECCV 2018", Munich, Germany, September 2018, https://hal.inria.fr/hal-01890980

[31] A. SOUA, O. SHAGDAR, J.-M. LASGOUTTES.*Toward Efficient Simulation Platform for Platoon Communication in Large Scale C-ITS Scenarios*, in "IEEE International Symposium on Networks, Computers and Communications", Roma, Italy, June 2018, https://hal.inria.fr/hal-01878153

[32] P. DE BEAUCORPS, A. VERROUST-BLONDET, R. PONCELET, F. NASHASHIBI.*RIS : A Framework for Motion Planning among Highly Dynamic Obstacles*, in "ICARCV 2018 - 15th International Conference on Control, Automation, Robotics and Vision", Singapour, Singapore, November 2018, https://hal.inria.fr/hal-01903318

### Conferences without Proceedings

[33] M. HADDED, P. MUHLETHALER, A. LAOUITI.*TDMA scheduling strategies for vehicular ad hoc networks: from a distributed to a centralized approach*, in "SoftCOM 2018 - 26th International Conference on Software, Telecommunications and Computer Networks", Split, Croatia, September 2018 [*DOI :* 10.23919/SOFTCOM.2018.8555781], https://hal.archives-ouvertes.fr/hal-01864550

[34] G. LE LANN.*Autonomic Vehicular Networks: Safety, Privacy, Cybersecurity and Societal Issues*, in "IEEE Vehicular Technology Conference Spring 2018 – First International Workshop on research advances in Cooperative ITS cyber security and privacy (C-ITSec)", Porto, Portugal, June 2018, https://arxiv.org/abs/1803.00424 , https://hal.archives-ouvertes.fr/hal-01720658

### Research Reports

[35] C. FURTLEHNER, J.-M. LASGOUTTES, A. ATTANASI, L. MESCHINI, M. PEZZULLA.*Spatio-temporal Probabilistic Short-term Forecasting on Urban Networks*, Inria Saclay, équipe TAU ; Inria de Paris, équipe RITS ; PTV-SISTeMA, December 2018, n$^{o}$ RR-9236, https://hal.inria.fr/hal-01964270

### Scientific Popularization

[36] R. DE CHARETTE, J. JONGWANE.*Les algorithmes de vision pour les véhicules autonomes*, in "Interstices", April 2018, https://hal.inria.fr/hal-01827600

### Other Publications

[37] E. NTARYAMIRA, C. MAXIM, C. FLORES, L. CUCU-GROSJEAN.*Towards temporal constraints in self driving cars*, July 2018, RTSOPS 2018 - 9th International Real-Time Scheduling Open Problems Seminar, Poster, https://hal.inria.fr/hal-01956016

[38] L. ROLDAO, R. DE CHARETTE, A. VERROUST-BLONDET.*A Statistical Update of Grid Representations from Range Sensors*, July 2018, https://arxiv.org/abs/1807.08483 - working paper or preprint, https://hal.inria.fr/hal-01847902

## References in notes

[39] G. FAYOLLE, C. FURTLEHNER.*About Hydrodynamic Limit of Some Exclusion Processes via Functional Integration*, in "Int. Math. Conf. “50 Years of IPPI”", Moscow, Institute for Information Transmission Problems (Russian Academy of Sciences), July 2011, Proceedings on CD. ISBN 978-5-901158-15-9, http://hal.inria.fr/hal-00662674

[40] G. FAYOLLE, R. IASNOGORODSKI.*Random Walks in the Quarter-Plane: Advances in Explicit Criterions for the Finiteness of the Associated Group in the Genus 1 Case*, in "Markov Processes and Related Fields", December 2015, vol. 21, n$^{o}$ 4, Accepted for publication in the journal MPRF (Markov Processes and Related Fields), https://hal.inria.fr/hal-01086684

[41] G. FAYOLLE, J.-M. LASGOUTTES.*Asymptotics and Scalings for Large Product-Form Networks via the Central Limit Theorem*, in "Markov Processes and Related Fields", 1996, vol. 2, n$^{o}$ 2, p. 317-348

[42] C. FURTLEHNER, Y. HAN, J.-M. LASGOUTTES, V. MARTIN, F. MARCHAL, F. MOUTARDE.*Spatial and Temporal Analysis of Traffic States on Large Scale Networks*, in "13th International IEEE Conference on Intelligent Transportation Systems ITSC'2010", Madère, Portugal, September 2010, https://hal-mines-paristech.archives-ouvertes.fr/hal-00527481

[43] G. LE LANN.*Safety in Vehicular Networks-On the Inevitability of Short-Range Directional Communications*, in "14th International Conference ADHOC-NOW, 2015", Athens, Greece, S. PAPAVASSILIOU, S. RUEHRUP (editors), Ad Hoc, Mobile, and Wireless Networks, Springer, June 2015, vol. Lecture Notes in Computer Science (LNCS), n$^{o}$ 9143, 14, Mobile Ad Hoc Networks [*DOI :* 10.1007/978-3-319-19662-6_24], https://hal.inria.fr/hal-01172595

[44] V. MARTIN, C. FURTLEHNER, Y. HAN, J.-M. LASGOUTTES.*GMRF Estimation under Topological and Spectral Constraints*, in "7th European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases", Nancy, France, T. CALDERS, F. ESPOSITO, E. HÜLLERMEIER, R. MEO (editors), Lecture Notes in Computer Science, Springer Berlin Heidelberg, September 2014, vol. 8725, p. 370-385 [*DOI :* 10.1007/978-3-662-44851-9_24], https://hal.archives-ouvertes.fr/hal-01065607

[45] V. MARTIN, J.-M. LASGOUTTES, C. FURTLEHNER.*Latent binary MRF for online reconstruction of large scale systems*, in "Annals of Mathematics and Artificial Intelligence", 2016, vol. 77, n$^{o}$ 1, p. 123-154

# Project-Team SECRET

## Security, Cryptology and Transmissions

RESEARCH CENTER
**Paris**

THEME
**Algorithmics, Computer Algebra and Cryptology**

# Table of contents

# Project-Team SECRET

*Creation of the Project-Team: 2008 July 01*

**Keywords:**

**Computer Science and Digital Science:**

A3.1.5. - Control access, privacy
A4. - Security and privacy
A4.2. - Correcting codes
A4.3. - Cryptography
A4.3.1. - Public key cryptography
A4.3.2. - Secret key cryptography
A4.3.3. - Cryptographic protocols
A4.3.4. - Quantum Cryptography
A7.1. - Algorithms
A7.1.4. - Quantum algorithms
A8.1. - Discrete mathematics, combinatorics
A8.6. - Information theory

**Other Research Topics and Application Domains:**

B6.4. - Internet of things
B6.5. - Information systems
B9.5.1. - Computer science
B9.5.2. - Mathematics
B9.10. - Privacy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Anne Canteaut [Team leader, Inria, Senior Researcher, HDR]
André Chailloux [Inria, Researcher]
Pascale Charpin [Inria, Emeritus, HDR]
Gaëtan Leurent [Inria, Starting Research Position until Feb. 2018, Researcher from March 2018]
Anthony Leverrier [Inria, Researcher, HDR]
María Naya Plasencia [Inria, Senior Researcher, HDR]
Nicolas Sendrier [Inria, Senior Researcher, HDR]
Jean-Pierre Tillich [Inria, Senior Researcher, HDR]

**Faculty Member**

Christina Boura [Univ de Versailles Saint-Quentin-en-Yvelines, Associate Professor, en délégation until Oct. 2018]

**PhD Students**

Xavier Bonnetain [Sorbonne Université]
Rémi Bricout [Sorbonne Université]
Rodolfo Canto Torres [Inria, until Oct 2018]
Kevin Carrier [Ministère de la Défense]
Daniel Coggia [DGA, from Sep 2018]

Thomas Debris [Sorbonne Université]
Sébastien Duval [Sorbonne Université, until Sep 2018]
Shouvik Ghorai [Sorbonne Université]
Antoine Grospellier [Sorbonne Université]
Matthieu Lequesne [Sorbonne Université]
Vivien Londe [Univ de Bordeaux]
Andrea Olivo [Inria]
Yann Rotella [Inria, until Sep 2018]
André Schrottenloher [Inria, from Feb 2018]
Ferdinand Sibleyras [Inria]
Valentin Vasseur [Univ René Descartes]

**Post-Doctoral Fellow**

Léo Perrin [Inria from Sept 2018, Fondation Sciences Mathématiques de Paris until Aug 2018]

**Visiting Scientists**

Thomas Peyrin [NTU, Singapore, January and July 2018]
Shizhu Tian [Univ. Chinese Academy of Sciences, from Sep 2018]

**Administrative Assistant**

Christelle Guiziou [Inria]

# 2. Overall Objectives

## 2.1. Presentation and scientific foundations

The research work within the project-team is mostly devoted to the design and analysis of cryptographic algorithms, in the classical or in the quantum setting. This work is essential since the current situation of cryptography is rather fragile. Many cryptographic protocols are now known whose security can be formally proved assuming that the involved cryptographic primitives are ideal (random oracle model, ideal cipher model...). However, the security of the available primitives has been either threatened by recent progress in cryptanalysis or by the possible invention of a large quantum computer. In other words, there is usually no concrete algorithm available to instantiate in practice the ideal "black boxes" used in these protocols!

In this context, our research work focuses on both families of cryptographic primitives, *symmetric* and *asymmetric* primitives.

## 2.2. Main topics

Our domain in cryptology includes the analysis and the design of

- symmetric primitives (a.k.a. secret-key algorithms),
- public-key primitives based on hard problems coming from coding theory which are likely to be resistant against a quantum computer,
- quantum cryptographic protocols whose security does not rely on computational assumptions but on the laws of quantum physics.

# 3. Research Program

## 3.1. Scientific foundations

Our approach relies on a competence whose impact is much wider than cryptology. Our tools come from information theory, discrete mathematics, probabilities, algorithmics, quantum physics... Most of our work mixes fundamental aspects (study of mathematical objects) and practical aspects (cryptanalysis, design of algorithms, implementations). Our research is mainly driven by the belief that discrete mathematics and algorithmics of finite structures form the scientific core of (algorithmic) data protection.

## 3.2. Symmetric cryptology

Symmetric techniques are widely used because they are the only ones that can achieve some major features such as high-speed or low-cost encryption, fast authentication, and efficient hashing. It is a very active research area which is stimulated by a pressing industrial demand. The process which has led to the new block cipher standard AES in 2001 was the outcome of a decade of research in symmetric cryptography, where new attacks have been proposed, analyzed and then thwarted by some appropriate designs. However, even if its security has not been challenged so far, it clearly appears that the AES cannot serve as a Swiss knife in all environments. In particular an important challenge raised by several new applications is the design of symmetric encryption schemes with some additional properties compared to the AES, either in terms of implementation performance (low-cost hardware implementation, low latency, resistance against side-channel attacks...) or in terms of functionalities (like authenticated encryption). The past decade has then been characterized by a multiplicity of new proposals. This proliferation of symmetric primitives has been amplified by several public competitions (eSTREAM, SHA-3, CAESAR...) which have encouraged innovative constructions and promising but unconventional designs. We are then facing up to a very new situation where implementers need to make informed choices among more than 40 lightweight block ciphers [0] or 57 new authenticated-encryption schemes [0]. Evaluating the security of all these proposals has then become a primordial task which requires the attention of the community.

In this context we believe that the cryptanalysis effort cannot scale up without an in-depth study of the involved algorithms. Indeed most attacks are described as ad-hoc techniques dedicated to a particular cipher. To determine whether they apply to some other primitives, it is then crucial to formalize them in a general setting. Our approach relies on the idea that a unified description of generic attacks (in the sense that they apply to a large class of primitives) is the only methodology for a precise evaluation of the resistance of all these new proposals, and of their security margins. In particular, such a work prevents misleading analyses based on wrong estimations of the complexity or on non-optimized algorithms. It also provides security criteria which enable designers to guarantee that their primitive resists some families of attacks. The main challenge is to provide a generic description which captures most possible optimizations of the attack.

## 3.3. Code-based cryptography

Public-key cryptography is one of the key tools for providing network security (SSL, e-commerce, e-banking...). The security of nearly all public-key schemes used today relies on the presumed difficulty of two problems, namely factorization of large integers or computing the discrete logarithm over various groups. The hardness of those problems was questioned in 1994 [0] when Shor showed that a quantum computer could solve them efficiently. Though large enough quantum computers that would be able to threaten the existing cryptosystems do not exist yet, the cryptographic research community has to get ready and has to prepare alternatives. This line of work is usually referred to as *post-quantum cryptography*. This has become a prominent research field. Most notably, an international call for post-quantum primitives [0] has been launched by the NIST, with a submission deadline in November 2017.

The research of the project-team in this field is focused on the design and cryptanalysis of cryptosystems making use of coding theory. Code-based cryptography is one the main techniques for post-quantum cryptography (together with lattice-based, multivariate, or hash-based cryptography).

## 3.4. Quantum information

The field of quantum information and computation aims at exploiting the laws of quantum physics to manipulate information in radically novel ways. There are two main applications:

[0] 35 are described on https://www.cryptolux.org/index.php/Lightweight_Block_Ciphers.
[0] see http://competitions.cr.yp.to/caesar-submissions.html
[0] P. Shor, *Algorithms for quantum computation: Discrete logarithms and factoring*, FOCS 1994.
[0] http://csrc.nist.gov/groups/ST/post-quantum-crypto/

- quantum computing, that offers the promise of solving some problems that seem to be intractable for classical computers such as for instance factorization or solving the discrete logarithm problem;
- quantum cryptography, which provides new ways to exchange data in a provably secure fashion. For instance it allows key distribution by using an authenticated channel and quantum communication over an unreliable channel with unconditional security, in the sense that its security can be proven rigorously by using only the laws of quantum physics, even with all-powerful adversaries.

Our team deals with quantum coding theoretic issues related to building a large quantum computer and with quantum cryptography. The first part builds upon our expertise in classical coding theory whereas the second axis focuses on obtaining security proofs for quantum protocols or on devising quantum cryptographic protocols (and more generally quantum protocols related to cryptography). A close relationship with partners working in the whole area of quantum information processing in the Parisian region has also been developed through our participation to the Fédération de Recherche "PCQC" (Paris Centre for Quantum Computing).

# 4. Application Domains

## 4.1. Cryptographic primitives

Our major application domain is the design of cryptographic primitives, especially for platforms with restricting implementation requirements. For instance, we aim at recommending (or designing) low-cost (or extremely fast) encryption schemes, or primitives which remain secure against quantum computers.

## 4.2. Code Reconstruction

To evaluate the quality of a cryptographic algorithm, it is usually assumed that its specifications are public, as, in accordance with Kerckhoffs principle, it would be dangerous to rely, even partially, on the fact that the adversary does not know those specifications. However, this fundamental rule does not mean that the specifications are known to the attacker. In practice, before mounting a cryptanalysis, it is necessary to strip off the data. This reverse-engineering process is often subtle, even when the data formatting is not concealed on purpose. A typical case is interception: some raw data, not necessarily encrypted, is observed out of a noisy channel. To access the information, the whole communication system has first to be disassembled and every constituent reconstructed. A transmission system actually corresponds to a succession of elements (symbol mapping, scrambler, channel encoder, interleaver... ), and there exist many possibilities for each of them. In addition to the "preliminary to cryptanalysis" aspect, there are other links between those problems and cryptology. They share some scientific tools (algorithmics, discrete mathematics, probability...), but beyond that, there are some very strong similarities in the techniques.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

- **Keynote at Eurocrypt:** A. Canteaut bas been an invited keynote speaker at Eurocrypt 2018 in Tel-Aviv.
- **Cryptanalysis of candidates to the NIST post-quantum competition:** The members of the project-team are involved in the design of several attacks against submissions to the NIST standardization effort for post-quantum cryptography. This work has led to the break of `EDON-K` key encapsulation mechanism, of `RLCE` encryption scheme, of `RankSign`, and of a recently proposed IBE scheme.

- **Quantum fault-tolerance with constant overhead:** In a couple of papers published at STOC 2018 and FOCS 2018, A. Grospellier and A. Leverrier together with O. Fawzi (from ENS Lyon) proved that quantum expander codes can be combined with quantum fault-tolerance techniques to achieve constant overhead: the ratio between the total number of physical qubits required for a quantum computation with faulty hardware and the number of logical qubits involved in the ideal computation is asymptotically constant, and can even be taken arbitrarily close to 1 in the limit of small physical error rate. This improves on the polylogarithmic overhead promised by the celebrated threshold theorem.

# 6. New Software and Platforms

## 6.1. CFS

FUNCTIONAL DESCRIPTION: Reference implementation of parallel CFS (reinforced version of the digital signature scheme CFS). Two variants are proposed, one with a « bit-packing » finite field arithmetic and an evolution with a « bit-slicing » finite-field arithmetic (collaboration with Peter Schwabe). For 80 bits of security the running time for producing one signature with the « bit-packing » variant is slightly above one second. This is high but was still the fastest so far. The evolution with the « bit-slicing » arithmetic produces the same signature in about 100 milliseconds.

- Participants: Grégory Landais and Nicolas Sendrier
- Contact: Nicolas Sendrier
- URL: https://gforge.inria.fr/projects/cfs-signature/

## 6.2. Collision Decoding

KEYWORDS: Algorithm - Binary linear code
FUNCTIONAL DESCRIPTION: Collision Decoding implements two variants of information set decoding : Stern-Dumer, and MMT. To our knowledge it is the best full-fledged open-source implementation of generic decoding of binary linear codes. It is the best generic attack against code-based cryptography.

- Participants: Grégory Landais and Nicolas Sendrier
- Contact: Nicolas Sendrier
- URL: https://gforge.inria.fr/projects/collision-dec/

## 6.3. ISDF

FUNCTIONAL DESCRIPTION: Implementation of the Stern-Dumer decoding algorithm, and of a varaint of the algorithm due to May, Meurer and Thomae.

- Participants: Grégory Landais and Nicolas Sendrier
- Contact: Anne Canteaut
- URL: https://gforge.inria.fr/projects/collision-dec/

# 7. New Results

## 7.1. Symmetric cryptology

**Participants:** Xavier Bonnetain, Christina Boura, Anne Canteaut, Pascale Charpin, Daniel Coggia, Sébastien Duval, Gaëtan Leurent, María Naya Plasencia, Léo Perrin, Yann Rotella, André Schrottenloher, Ferdinand Sibleyras.

### *7.1.1. Block ciphers*

Our recent results mainly concern either the analysis or the design of lightweight block ciphers.

**Recent results:**

- Nonlinear approximations of block ciphers: A. Canteaut, together with C. Beierle and G. Leander have exhibited the relationship between nonlinear invariants for block ciphers and nonlinear approximations. They have shown that, in some cases, the linear hull effect may be formalized in terms of nonlinear invariants. They have also introduced a new framework to study the probability of nonlinear approximations over iterated block ciphers [13], [26]
- Impossible differential cryptanalysis: C. Boura, V. Lallemand and M. Naya-Plasencia have introduced new techniques and complexity analyses for impossible differential cryptanalysis. They also showed that the technique of multiple differentials can be applied to impossible differential attacks [16]
- Construction of lightweight MDS matrices: S. Duval and G. Leurent have exhibited MDS matrices with the lowest known implementation cost. They have been constructed by a search through a space of circuits yielding MDS matrices [20], [11]

### *7.1.2. Stream ciphers*

Stream ciphers provide an alternative to block-cipher-based encryption schemes. They are especially well-suited in applications which require either extremely fast encryption or a very low-cost hardware implementation.

**Recent results:**

- Design of encryption schemes for efficient homomorphic-ciphertext compression: A. Canteaut, M. Naya-Plasencia together with their coauthors have investigated the constraints on the symmetric cipher imposed by this application and they have proposed some solutions based on additive IV-based stream ciphers [17].
- Cryptanalysis of Goldreich pseudo-random generator: Goldreich's PRG is a theoretical construction which expands a short random string into a long pseudo-random string by applying a simple $d$-ary predicate to public random sized subsets of the bits of the seed. While the security of Goldreich's PRG has been thoroughly investigated, with a variety of results deriving provable security guarantees against classes of attacks in some parameter regimes and necessary criteria to be satisfied by the underlying predicate, little was known about its concrete security and efficiency. Motivated by the hope of getting practical instantiations of this construction, Y. Rotella and his co-authors initiated a study of the concrete security of Goldreich's PRG, and evaluated its resistance to cryptanalytic attacks. They developped a new guess-and-determine-style attack, and identified new criteria which captured the security guarantees [44].

### *7.1.3. Authenticated encryption*

A limitation of all classical block ciphers is that they aim at protecting confidentiality only, while most applications need both encryption and authentication. These two functionalities are provided by using a block cipher like the AES together with an appropriate mode of operation. However, it appears that the most widely-used mode of operation for authenticated encryption, AES-GCM, is not very efficient for high-speed networks. Also, the security of the GCM mode completely collapses when an IV is reused. These severe drawbacks have then motivated an international competition named CAESAR, partly supported by the NIST, which has been launched in order to define some new authenticated encryption schemes [0]. The project-team is involved in a national cryptanalytic effort in this area led by the BRUTUS project funded by the ANR. In this context, the members of the project-team have obtained some cryptanalytic results on several candidates to the CAESAR competition.

**Recent results:**

- State-recovery attack on a simplified version of Ketje Jr. [21], [34]
- Cryptanalysis of Morus, one of the finalists of the CAESAR competition [42]

[0] http://competitions.cr.yp.to/caesar.html

### *7.1.4. Cryptographic properties and construction of appropriate building blocks*

The construction of building blocks which guarantee a high resistance against the known attacks is a major topic within our project-team, for stream ciphers, block ciphers and hash functions. The use of such optimal objects actually leads to some mathematical structures which may be at the origin of new attacks. This work involves fundamental aspects related to discrete mathematics, cryptanalysis and implementation aspects. Actually, characterizing the structures of the building blocks which are optimal regarding to some attacks is very important for finding appropriate constructions and also for determining whether the underlying structure induces some weaknesses or not. For these reasons, we have investigated several families of filtering functions and of S-boxes which are well-suited for their cryptographic properties or for their implementation characteristics.

**Recent results:**

- Differential Equivalence of Sboxes: C. Boura, A. Canteaut and their co-authors have studied two notions of differential equivalence of Sboxes corresponding to the case when the functions have the same difference table, or when their difference tables have the same support [15], [25]. They proved that these two notions do not coincide, and that they are invariant under some classical equivalence relations like EA and CCZ equivalence. They also proposed an algorithm for determining the whole equivalence class of a given function.
- Boomerang Uniformity of Sboxes: The boomerang attack is a cryptanalysis technique against block ciphers which combines two differentials for the upper part and the lower part of the cipher. The Boomerang Connectivity Table (BCT) is a tool introduced by Cid *et al.* at Eurocrypt 2018 for analysing the dependency between these two differentials. C. Boura and A. Canteaut [14] have provided an in-depth analysis of BCT, by studying more closely differentially 4-uniform Sboxes. They have completely characterized the BCT of all differentially 4-uniform permutations of 4 bits and then study these objects for some cryptographically relevant families of Sboxes, as the inverse function and quadratic permutations. These two families are the first examples of differentially 4-uniform Sboxes optimal against boomerang attacks for an even number of variables, answering an open question raised by Cid *et al.*.
- CCZ equivalence of Sboxes: A. Canteaut and L. Perrin have characterized CCZ-equivalence as a property of the zeroes in the Walsh spectrum of an Sbox (or equivalently in their DDT). They used this framework to show how to efficiently upper bound the number of distinct EA-equivalence classes in a given CCZ-equivalence class. More importantly, they proved that CCZ-equivalence can be reduced to the association of EA-equivalence and an operation called twisting. They then revisited several results from the literature on CCZ-equivalence and showed how they can be interpreted in light of this new framework [18], [29]
- Links between linear and differential properties of Sboxes: P. Charpin together with J. Peng has established new links between the differential uniformity and the nonlinearity of some Sboxes in the case of two-valued functions and quadratic functions. More precisely, they have exhibited a lower bound on the nonlinearity of monomial permutations depending on their differential uniformity, as well as an upper bound in the case of differentially two-valued functions [19], [55]
- Construction of building-blocks with resistance against fault-attacks at a low implementation overhead [50].

### *7.1.5. Modes of operation and generic attacks*

In order to use a block cipher in practice, and to achieve a given security notion, a mode of operation must be used on top of the block cipher. Modes of operation are usually studied through provable security, and we know that their use is secure as long as the underlying primitive is secure, and we respect some limits on the amount of data processed. The analysis of generic attack helps us understand what happens when the hypotheses of the security proofs do not hold, or the corresponding limits are not respected. Comparing proofs and attacks also shows gaps where our analysis is incomplete, and when improved proof or attacks are required.

**Recent results:**

- Use of block ciphers operating on small blocks with the CTR mode [53]: the security proof of the CTR mode requires that no more than $2^{n/2}$ blocks are encrypted with the same key, but the known attacks reveal very little information and are considered less problematic than on CBC. However, G. Leurent and F. Sibleyras have exhibited concrete attacks against the CTR mode when processing close to $2^{n/2}$ blocks of data, and have shown that an attacker can actually extract as much information as in the case of CBC encryption.
- Generic attacks against some MAC constructions based on block ciphers [52]: G. Leurent and F. Sibleyras, together with M. Nandi, have studied the security of several recent MAC constructions with provable security beyond the birthday bound, namely `SUM-ECBC`, `PMAC+`, `3kf9`, `GCM-SIV2`, and some variants. They described a new cryptanalysis technique for double-block MACs and they showed how to build a forgery attack with query complexity $\mathcal{O}(2^{3n/4})$, proving that these schemes do not reach full security in the information-theoretic model. Surprisingly, their attack on `LightMAC+` invalidates a recent security proof by Naito. Moreover, they gave the first attack against `SUM-ECBC` and `GCM-SIV2`, with complexity below $2^n$.

## 7.2. Code-based cryptography

**Participants:** Rodolfo Canto Torres, Thomas Debris, Matthieu Lequesne, Nicolas Sendrier, Jean-Pierre Tillich, Valentin Vasseur.

The first cryptosystem based on error-correcting codes was a public-key encryption scheme proposed by McEliece in 1978; a dual variant was proposed in 1986 by Niederreiter. We proposed the first (and only) digital signature scheme in 2001. Those systems enjoy very interesting features (fast encryption/decryption, short signature, good security reduction) but also have their drawbacks (large public key, encryption overhead, expensive signature generation). Some of the main issues in this field are

- security analysis, including against a quantum adversary, implementation and practicality of existing solutions,
- reducing the key size, *e.g.*, by using rank metric instead of Hamming metric, or by using structured codes,
- addressing new functionalities, like identity-based encryption, hashing or symmetric encryption.

Our recent work on code-based cryptography has to be seen in the context of the recently launched NIST competition whose purpose is to standardize quantum-safe public-key primitives. This call concerns all three major cryptographic primitives, namely public-key cryptosytems, key-exchange protocols and digital signature schemes. The most promising techniques today for addressing this issue are code-based cryptography, lattice-based cryptography, mutivariate cryptography, and hash-based cryptography.

Our contributions in this area are two-fold and consist in:

- designing and analysis new code-based solutions;
- cryptanalyzing code-based schemes, especially candidates to the NIST competition.

### *7.2.1. Design of new code-based solutions*

The members of the project-team have submitted several candidates to the NIST competition, including a key-exchange protocol based on quasi-cyclic MDPC codes [41]. Their recent work on MDPC codes is important in this context in order to carefully analyze the properties of this candidate.

**Recent results:**

- Thwarting the GJS attack: the decryption algorithm of the QC-MDPC cryptosystem is based on an iterative bit-flipping algorithm, which fails with a small probability. These failures have been exploited in 2016 by Guo, Johansson and Stankovski to perform a key-recovery attack. JP Tillich recently analyzed how this attack can be avoided by increasing the key size of the scheme. Most notably, he proved that, under a very reasonable assumption, the error probability after decoding decays almost exponentially with the code-length with just two iterations of bit-flipping. With an additional assumption, it even decays exponentially with an unbounded number of iterations, implying that in this case the increase of the key size equired for resisting to the GJS attack is only moderate [54].
- Design of a new KEM with IND-CCA2 security in a model considering decoding failures [46]: M. Lequesne, N. Sendrier and their co-authors explored the underlying causes of the GJS attack, how it can be improved and how it can be mitigated. They derived a new timing attack performing well even in cases which were more challenging to the GJS attack. They also showed how to construct a new KEM, called ParQ that can reduce the decryption failure rate to a level negligible in the security parameter. They formally proved the IND-CCA2 security of ParQ, in a model that considers decoding failures.
- Design of a new code-based signature scheme [81]: T. Debris, N. Sendrier and JP Tillich recently proposed a "hash-and-sign" code-based signature scheme called `Wave`, which uses a family of ternary generalized (U, U + V) codes. `Wave` achieves existential unforgeability under adaptive-chosen-message attacks in the random oracle model with a tight reduction to two assumptions from coding theory: one is a distinguishing problem that is related to the trapdoor inserted in the scheme, the other one is a multiple-target version of syndrome decoding. This scheme enjoys efficient signature and verification algorithms. For 128-bit security, signature are 8000-bit long and the public-key size is slightly smaller than one megabyte.

### *7.2.2. Cryptanalysis of code-based schemes*

**Recent results:**

- Cryptanalysis of two public-key cryptosystems based on the rank syndrome decoding problem [41]: JP Tillich and his co-authors proposed an attack on the Rank Syndrome Decoding problem which improves the previously best known algorithm for solving this problem. This attack breaks for some parameters some recently proposed cryptosystems based on LRPC codes or Gabidulin codes, including Loidreau's cryptosystem and the LRPC cryptosystem.
- Cryptanalysis of the NIST submission `RankSign` and of a recently proposed IBE scheme: T. Debris and JP Tillich have presented an algebraic attack against `RankSign` that exploits the fact that the augmented LRPC codes used in this scheme have codewords with a very low weight. This attack shows that all the parameters proposed for this candidate can be broken. They also proved that, for the IBE scheme based on `RankSign`, the problem is deeper than finding a new signature in rank-based cryptography, since they found an attack on the generic problem upon which the security reduction relies [45].
- Cryptanalysis of the `EDON-K` key encapsulation mechanism submitted to the NIST competition: `EDON-K` is a candidate to the NIST competition which is inspired by the McEliece scheme but uses another family of codes defined over $\mathbb{F}_{2^{128}}$ instead of $\mathbb{F}_2$ and is not based on the Hamming metric. M. Lequesne and JP Tillich presented an attack making the scheme insecure for the intended use. Indeed, recovering the error in the McEliece scheme corresponding to `EDON-K` can be viewed as a decoding problem for the rank-metric with a super-code of an LRPC code of very small rank A suitable parity-check matrix for this super-code can then be easily derived from the public key and used to recover the error [51].

- Attack against RLCE [80]: M. Lequesne and JP Tillich, together with A. Couvreur, recently presented a key-recovery attack against the Random Linear Code Encryption (RLCE) scheme recently submitted by Y. Wang to the NIST competition. This attack recovers the secret-key for all the short key-parameters proposed by the author. It uses a polynomial-time algorithm based on a square code distinguisher.

## 7.3. Quantum Information

**Participants:** Xavier Bonnetain, Rémi Bricout, André Chailloux, Shouvik Ghorai, Antoine Grospellier, Anirudh Krishna, Anthony Leverrier, Vivien Londe, María Naya Plasencia, Andrea Olivo, Jean-Pierre Tillich, André Schrottenloher.

Our research in quantum information focusses on several axes: quantum codes with the goal of developing better error correction strategies to build large quantum computers, quantum cryptography which exploits the laws of quantum mechanics to derive security guarantees, relativistic cryptography which exploits in addition the fact that no information can travel faster than the speed of light and finally quantum cryptanalysis which investigates how quantum computers could be harnessed to attack classical cryptosystems.

### *7.3.1. Quantum codes*

Protecting quantum information from external noise is an issue of paramount importance for building a quantum computer. It also worthwhile to notice that all quantum error-correcting code schemes proposed up to now suffer from the very same problem that the first (classical) error-correcting codes had: there are constructions of good quantum codes, but for the best of them it is not known how to decode them in polynomial time.

Two PhD students within the project-team work on this topic. First, Antoine Grospellier, co-advised by A. Leverrier and O. Fawzi (Ens Lyon), studies efficient decoding algorithms for quantum LDPC codes. Beyond their intrinsic interest for channel-coding problems, such algorithms would be particularly relevant in the context of quantum fault-tolerance, since they would allow to considerably reduce the required overhead to obtain fault-tolerance in quantum computation. Vivien Londe is co-advised by A. Leverrier and G. Zémor (IMB) and his thesis is devoted to the design of better quantum LDPC codes: the main idea is to generalize the celebrated toric code of Kitaev by considering cellulations of manifolds in higher dimensions. A recent surprising result was that this approach leads to a much better behaviour than naively expected and a major challenge is to explore the mathematics behind this phenomenon in order to find even better constructions, or to uncover potential obstructions.

**Recent results:**

- Decoding algorithm for quantum expander codes [48], [47], [56] In this work, A. Grospellier, A. Leverrier and O. Fawzi analyze an efficient decoding algorithm for quantum expander codes and prove that it can correct a linear number of random errors with a negligible failure probability. As an application, this shows that this family of codes can be used to obtain quantum fault-tolerance with only a constant overhead in terms of qubits, compared to a polylogarithmic overhead as in previous schemes. This is a crucial step in order to eventually build large universal quantum computers.

### *7.3.2. Quantum cryptography*

Quantum cryptography exploits the laws of quantum physics to establish the security of certain cryptographic primitives. The most studied one is certainly quantum key distribution, which allows two distant parties to establish a secret using an untrusted quantum channel. Our activity in this field is particularly focussed on protocols with continuous variables, which are well-suited to implementations. The interest of continuous variables for quantum cryptography was recently recognized by being awarded a 10 M€ funding from the Quantum Flagship and SECRET will contribute to this project by studying the security of new key distribution protocols [88].

**Recent results:**

- Security proof for two-way continuous-variable quantum key distribution [22]: while many quantum key distribution protocols are one-way in the sense that quantum information is sent from one party to the other, it can be beneficial in terms of performance to consider two-way protocols where the quantum states perform a round-trip between the two parties. In this paper (to appear in *Physical Review A*), we show how to exploit the symmetries of the protocols in phase-space to establish their security against the most general attacks allowed by quantum theory.
- Investigating the optimality of ancilla-assisted linear optical Bell measurements [24]: Due to its experimental and theoretical simplicity, linear quantum optics has proved to be a promising route for the early implementation of important quantum communication protocols. A. Olivo and F. Grosshans study the efficiency of non ambiguous Bell measurements in this model and show both theoretical and numerical bounds depending on the number of ancilla qubits. This is important for understanding what resources are needed for building quantum repeaters, the last missing building block for secure long distance quantum key distribution networks.

### *7.3.3. Relativistic cryptography*

Two-party cryptographic tasks are well-known to be impossible without complexity assumptions, either in the classical or the quantum world. Remarkably, such no-go theorems become invalid when adding the physical assumption that no information can travel faster than the speed of light. This additional assumption gives rise to the emerging field of relativistic cryptography. We worked on this topic for several years and Andrea Olivo was recruited as a PhD student to continue working on both theoretical and practical aspects of relativistic cryptography.

**Recent results:**

- Relativistic commitment and zero-knowledge proofs [30]: A. Chailloux and A. Leverrier constructed a relativistic zero-knowledge protocol for any NP-complete problem. The main technical tool is the analysis of quantum consecutive measurements, which allows us to prove security against quantum adversaries. R. Bricout and A. Chailloux also studied relativistic multi-round bit commitment schemes. They showed optimal classical cheating strategies for the canonical $F_Q$ commitment scheme.

### *7.3.4. Quantum cryptanalysis of symmetric primitives*

Symmetric cryptography seems at first sight much less affected in the post-quantum world than asymmetric cryptography: its main known threat seemed for a long time Grover's algorithm, which allows for an exhaustive key search in the square root of the normal complexity. For this reason, it was usually believed that doubling key lengths suffices to maintain an equivalent security in the post-quantum world. However, a lot of work is certainly required in the field of symmetric cryptography in order to "quantize" the classical families of attacks in an optimized way, as well as to find new dedicated quantum attacks. M. Naya Plasencia has recently been awarded an ERC Starting grant for her project named QUASYModo on this topic.

**Recent results:**

- Hidden-shift quantum cryptanalysis [43]: X. Bonnetain and M. Naya-Plasencia have obtained new results that consider the tweak proposed at Eurocrypt 2017 of using modular additions to counter Simon's attacks. They have developed new algorithms that improve and generalize Kuperberg's algorithm for the hidden shift problem. Thanks to their improved algorithm, they have been able to build a quantum attack in the superposition model on Poly1305, proposed at FSE 2005, largely used and claimed to be quantumly secure. They also analyzed the security of some classical symmetric constructions with concrete parameters, to evaluate the impact and practicality of the proposed tweak, concluding that it does not seem to be efficient
- Quantum algorithm for the $k$-XOR problem [49]: The $k$-XOR (or generalized birthday) problem aims at finding $k$ elements of $n$-bits, drawn at random, such that the XOR of all of them is 0. The algorithms proposed by Wagner more than 15 years ago remain the best known classical algorithms

for solving it, when disregarding logarithmic factors. M. Naya-Plasencia and A. Schrottenloher, together with L. Grassi, studied this problem in the quantum setting and provided algorithms with the best known quantum time-complexities. In particular, they were able to considerably improve the 3-XOR algorithm.

- Quantum cryptanalysis of CSIDH and Ordinary Isogeny-based Schemes [68]: CSIDH is a recent proposal by Castryck et al. for post-quantum non-interactive key-exchange. It is similar in design to a scheme by Couveignes, Rostovtsev and Stolbunov, but it replaces ordinary elliptic curves by supersingular elliptic curves. Although CSIDH uses supersingular curves, it can attacked by a quantum subexponential hidden shift algorithm due to Childs et al. While the designers of CSIDH claimed that the parameters they suggested ensures security against this algorithm, X. Bonnetain and A. Schrottenloher showed that these security parameters were too optimistic: they improved the hidden shift algorithm and gave a precise complexity analysis in this context, which greatly reduced the complexity. For example, they showed that only $2^{35}$ quantum equivalents of a key-exchange are sufficient to break the 128-bit classical, 64-bit quantum security parameters proposed, instead of $2^{62}$. They also extended their analysis to ordinary isogeny computations, and showed that an instance proposed by De Feo, Kieffer and Smith and expected to offer 56 bits of quantum security can be broken in $2^{38}$ quantum evaluations of a key exchange.

# 8. Partnerships and Cooperations

## 8.1. National Initiatives

### *8.1.1. ANR*

- **ANR BRUTUS** ($10/14 \rightarrow 09/18$)
  *Authenticated Ciphers and Resistance against Side-Channel Attacks*
  ANR program: Défi Société de l'information et de la communication
  Partners: ANSSI, Inria (project-team SECRET and project-team MARELLE), Orange, University of Lille, University of Rennes, University Versailles-Saint Quentin
  160 kEuros
  The Brutus project aims at investigating the security of authenticated encryption systems. We plan to evaluate carefully the security of the most promising candidates to the CAESAR competition, by trying to attack the underlying primitives or to build security proofs of modes of operation. We target the traditional black-box setting, but also more "hostile" environments, including the hardware platforms where some side-channel information is available.
- **ANR DEREC** ($10/16 \rightarrow 09/21$)
  *Relativistic cryptography*
  ANR Program: jeunes chercheurs
  244 kEuros
  The goal of project DEREC is to demonstrate the feasibility of guaranteeing the security of some cryptographic protocols using the relativistic paradigm, which states that information propagation is limited by the speed of light. We plan to study some two party primitives such as bit commitment and their security against classical and quantum adversaries in this model. We then plan to the integration of those primitives into larger cryptosystems. Finally, we plan on performing a demonstration of those systems in real life conditions.

- **ANR CBCRYPT** ($10/17 \rightarrow 09/21$)
  *Code-based cryptography*
  ANR Program: AAP Générique 2017
  Partners: Inria SECRET (coordinator), XLIM, Univ. Rouen, Univ. Bordeaux.
  197 kEuros
  The goal of CBCRYPT is to propose code-based candidates to the NIST call aiming at standardizing public-key primitives which resist to quantum attacks. These proposals are based either on code-based schemes relying on the usual Hamming metric or on the rank metric. The project does not deal solely with the NIST call. We also develop some other code-based solutions: these are either primitives that are not mature enough to be proposed in the first NIST call or whose functionalities are not covered by the NIST call, such as identity-based encryption, broadcast encryption, attribute based encryption or functional encryption. A third goal of this project is of a more fundamental nature: namely to lay firm foundations for code-based cryptography by developing thorough and rigorous security proofs together with a set of algorithmic tools for assessing the security of code-based cryptography.
- **ANR quBIC** ($10/17 \rightarrow 09/21$)
  *Quantum Banknotes and Information-Theoretic Credit Cards*
  ANR Program: AAP Générique 2017
  Partners: Univ. Paris-Diderot (coordinator), Inria SECRET, UPMC (LIP6), CNRS (Laboratoire Kastler Brossel)
  87 kEuros
  For a quantum-safe future, classical security systems as well as quantum protocols that guarantee security against all adversaries must be deployed. Here, we will study and implement one of the most promising quantum applications, namely unforgeable quantum money. A money scheme enables a secure transaction between a client, a vendor and a bank via the use of a credit card or via the use of banknotes, with maximal security guarantees. Our objectives are to perform a theoretical analysis of quantum money schemes, in realistic conditions and for encodings in both discrete and continuous variables, and to demonstrate experimentally these protocols using state-of-the-art quantum memories and integrated detection devices.

# 8.2. European Initiatives

## *8.2.1. FP7 & H2020 Projects*

### *8.2.1.1. PQCRYPTO*

Title: Post-quantum cryptography for long-term security

Programm: H2020

Duration: March 2015 - March 2018

Coordinator: TECHNISCHE UNIVERSITEIT EINDHOVEN

Partners:

Academia Sinica (Taiwan)

Bundesdruckerei (Germany)

Danmarks Tekniske Universitet (Denmark)

Katholieke Universiteit Leuven (Belgium)

Nxp Semiconductors Belgium Nv (Belgium)

Ruhr-Universitaet Bochum (Germany)

Stichting Katholieke Universiteit (Netherlands)

Technische Universiteit Eindhoven (Netherlands)

Technische Universitaet Darmstadt (Germany)

University of Haifa (Israel)

Inria contact: Nicolas Sendrier

Online banking, e-commerce, telemedicine, mobile communication, and cloud computing depend fundamentally on the security of the underlying cryptographic algorithms. Public-key algorithms are particularly crucial since they provide digital signatures and establish secure communication without requiring in-person meetings. Essentially all applications today are based on RSA or on the discrete-logarithm problem in finite fields or on elliptic curves. Cryptographers optimize parameter choices and implementation details for these systems and build protocols on top of these systems; cryptanalysts fine-tune attacks and establish exact security levels for these systems. Alternative systems are far less visible in research and unheard of in practice. It might seem that having three systems offers enough variation, but these systems are all broken as soon as large quantum computers are built. The EU and governments around the world are investing heavily in building quantum computers; society needs to be prepared for the consequences, including cryptanalytic attacks accelerated by these computers. Long-term confidential documents such as patient health-care records and state secrets have to guarantee security for many years, but information encrypted today using RSA or elliptic curves and stored until quantum computers are available will then be as easy to decipher as Enigma-encrypted messages are today. PQCRYPTO will allow users to switch to post-quantum cryptography: cryptographic systems that are not merely secure for today but that will also remain secure long-term against attacks by quantum computers. PQCRYPTO will design a portfolio of high-security post-quantum public-key systems, and will improve the speed of these systems, adapting to the different performance challenges of mobile devices, the cloud, and the Internet of Things. PQCRYPTO will provide efficient implementations of high-security post-quantum cryptography for a broad spectrum of real-world applications.

*8.2.1.2. QCALL*

Title: Quantum Communications for ALL

Programm: H2020-MSCA-ITN-2015

Duration: December 2016 - November 2020

Coordinator: University of Leeds (UK)

Other partners: see http://www.qcall-itn.eu/

Inria contact: Anthony Leverrier

QCALL is a European Innovative Training Network that endeavors to take the next necessary steps to bring the developing quantum technologies closer to the doorsteps of end users. QCALL will empower a nucleus of 15 doctoral researchers in this area to provide secure communications in the European continent and, in the long run, to its connections worldwide.

*8.2.1.3. ERC QUASYModo*

Title: QUASYModo *Symmetric Cryptography in the Post-Quantum World*

Program: ERC starting grant

Duration: September 2017 - August 2022

PI: María Naya Plasencia

As years go by, the existence of quantum computers becomes more tangible and the scientific community is already anticipating the enormous consequences of the induced breakthrough in computational power. Cryptology is one of the affected disciplines. Indeed, the current state-of-the-art asymmetric cryptography would become insecure, and we are actively searching for alternatives. Symmetric cryptography, essential for enabling secure communications, seems much less affected at first sight: its biggest known threat is Grover's algorithm, which allows exhaustive key searches in the square root of the normal complexity. Thus, so far, it is believed that doubling key lengths suffices to maintain an equivalent security in the post- quantum world. The security of symmetric cryptography is completely based on cryptanalysis: we only gain confidence in the security of a symmetric

primitive through extensive and continuous scrutiny. It is therefore not possible to determine whether a symmetric primitive might be secure or not in a post-quantum world without first understanding how a quantum adversary could attack it. Correctly evaluating the security of symmetric primitives in the post-quantum world cannot be done without a corresponding cryptanalysis toolbox, which neither exists nor has ever been studied. This is the big gap I have identified and that I plan to fill with this project. Next, doubling the key length is not a trivial task and needs to be carefully studied. My ultimate aim is to propose efficient solutions secure in the post-quantum world with the help of our previously obtained quantum symmetric cryptanalysis toolbox. This will help prevent the chaos that big quantum computers would generate: being ready in advance will definitely save a great amount of time and money, while protecting our current and future communications. The main challenge of QUASYModo is to redesign symmetric cryptography for the post-quantum world.

### *8.2.2. Collaborations in European Programs, Except FP7 & H2020*

*8.2.2.1. QCDA*

Program: QuantERA ERA-NET Cofund in Quantum Technologies

Project acronym: QCDA

Project title: Quantum Code Design and Architecture

Duration: February 2018 - January 2021

Coordinator: Earl Campbell, University of Sheffield, UK

Other partners: University of Sheffield (UK), TU Delft (Netherlands), TU Munich (Germany), University College London (UK)

Inria contact: Anthony Leverrier

General purpose quantum computers must follow a fault-tolerant design to prevent ubiquitous decoherence processes from corrupting computations. All approaches to fault-tolerance demand extra physical hardware to perform a quantum computation. Kitaev's surface, or toric, code is a popular idea that has captured the hearts and minds of many hardware developers, and has given many people hope that fault-tolerant quantum computation is a realistic prospect. Major industrial hardware developers include Google, IBM, and Intel. They are all currently working toward a fault-tolerant architecture based on the surface code. Unfortunately, however, detailed resource analysis points towards substantial hardware requirements using this approach, possibly millions of qubits for commercial applications. Therefore, improvements to fault-tolerant designs are a pressing near-future issue. This is particularly crucial since sufficient time is required for hardware developers to react and adjust course accordingly.

This consortium will initiate a European co-ordinated approach to designing a new generation of codes and protocols for fault-tolerant quantum computation. The ultimate goal is the development of high-performance architectures for quantum computers that offer significant reductions in hardware requirements; hence accelerating the transition of quantum computing from academia to industry. Key directions developed to achieve these improvements include: the economies of scale offered by large blocks of logical qubits in high-rate codes; and the exploitation of continuous-variable degrees of freedom.

The project further aims to build a European community addressing these architectural issues, so that a productive feedback cycle between theory and experiment can continue beyond the lifetime of the project itself. Practical protocols and recipes resulting from this project are anticipated to become part of the standard arsenal for building scalable quantum information processors.

## 8.3. International Initiatives

### *8.3.1. Inria Associate Teams Not Involved in an Inria International Labs*

*8.3.1.1. CHOCOLAT*

Title: Chosen-prefix Collision Attack on SHA-1 with ASICs Cluster

International Partner (Institution - Laboratory - Researcher):

NTU (Singapore) - SYLLAB - Peyrin Thomas

Start year: 2017

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

The hash function SHA-1 is one of the most widely used hash functions in the industry, but it has been shown to not be collision-resistant by a team of Chinese researchers led by Prof. Wang in 2005. However, nobody has publicly produced a real pair of colliding messages so far, because the estimated attack complexity is around $2^{63}$ SHA-1 computations (this represents about 70000 years of computation on a normal PC).

While a collision of SHA-1 would clearly demonstrate the weakness of the algorithm, a much more powerful attack would be to find a collision such that the prefix of the colliding messages is chosen by some challenger beforehand. In particular, this would allow creating a rogue certificate authority certificate that would be accepted by browsers. Such an attack has already been deployed for certificates using the MD5 hash function, but MD5 is much weaker than SHA-1 and it has already been removed from most security applications. SHA-1 is still widely used and performing such an attack for certificates using SHA-1 would have a very big impact.

The objective of the project is to design a chosen-prefix collision attack against the SHA-1 hash function, and to implement the attack in practice. We estimate this will require $2^{70}$ computations, and we will use an ASIC cluster to perform such a computation.

### 8.3.2. Inria International Partners

#### 8.3.2.1. Declared Inria International Partners

Title: Discrete Mathematics, Codes and Cryptography

International Partner (Institution - Laboratory - Researcher):

Indian Statistical Institute (India) - Cryptology Research Group - Bimal Roy

Duration: 2014 - 2018

Start year: 2014

Today's cryptology offers important challenges. Some are well-known: Can we understand existing cryptanalysis techniques well enough to devise criterion for the design of efficient and secure symmetric cryptographic primitives? Can we propose cryptographic protocols which offer provable security features under some reasonable algorithmic assumptions? Some are newer: How could we overcome the possible apparition of a quantum computer with its devastating consequences on public key cryptography as it is used today? Those challenges must be addressed, and some of the answers will involve tools borrowed to discrete mathematics, combinatorics, algebraic coding theory, algorithmic. The guideline of this proposal is to explore further and enrich the already well established connections between those scientific domains and their applications to cryptography and its challenges.

#### 8.3.2.2. Informal International Partners

- Nanyang Technological University (Singapore): cryptanalysis of symmetric primitives.
- Ruhr-Universität Bochum (Germany): design and cryptanalysis of symmetric primitives.
- University of Sherbrooke (Canada): quantum codes.

## 8.4. International Research Visitors

### 8.4.1. Visits of International Scientists

- Thomas Peyrin, NTU Singapore, January 2018 and July 2018.

- Sristy Agrawal, Indian Institute of Science Education and Research, Kolkata, India, January 2018.
- Anastasiya Gorodilova, Sobolev Institute of Mathematics, Novosibirsk, Russia, September 2018.
- Lorenzo Grassi, IAIK, Graz University of Technology, Austria, September 2018.

*8.4.1.1. Internships*

- Daniel Coggia, MPRI, March-Aug. 2018
- Anaïs Querol Cruz, MPRI, March-Aug. 2018
- Florian Wartelle, UVSQ, March-Sept. 2018

### *8.4.2. Visits to International Teams*

*8.4.2.1. Research Stays Abroad*

- NTU, Singapore, joint work within the CHOCOLAT Associate Team: S. Duval (April 8-19), G. Leurent (October 29 - November 10).
- University of Sherbrooke, Sherbrooke, Canada, June 11-15, 2018 (J.P. Tillich)
- Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, September 30-October 9, 2018 (P. Charpin).

# 9. Dissemination

## 9.1. Promoting Scientific Activities

### *9.1.1. Scientific Events Organisation*

*9.1.1.1. General Chair, Scientific Chair*

- WCC 2019, March 31 - April 5, 2019, St Jacut-de-la-Mer, France: A. Canteaut, program co-chair
- Eurocrypt 2020, Zagreb, Croatia: A. Canteaut, program co-chair
- Workshop on quantum code design and architectures (kick-off meeting of the European project QCDA), November 5-6, 2018, Paris (France): A. Leverrier.

*9.1.1.2. Member of the Organizing Committees*

- Training School on Symmetric Cryptography and Blockchain: February 19-23, 2018, Torremolinos (Spain): A. Canteaut.

### *9.1.2. Scientific Events Selection*

*9.1.2.1. Chair of Conference Program Committees*

As a co-editor-in-chief of the journal *IACR Transactions on Symmetric Cryptology*, María Naya-Plasencia served as a program chair of the conference *Fast Software Encryption (FSE)*, held in Bruges March 2018. Gaëtan Leurent will serve as a co-editor-in-chief of *IACR Transactions on Symmetric Cryptology* starting from 2019.

*9.1.2.2. Member of the Conference Program Committees*

- FSE 2018: March 5-7, 2018, Bruges, Belgium (C. Boura, A. Canteaut, G. Leurent, M. Naya-Plasencia, L. Perrin);
- CryptoAction Symposium 2018: April 4-5, Sutomore, Montenegro (A. Canteaut);
- PQCrypto 2018: April 9-11, 2018, Fort Lauderdale, USA, (M. Naya-Plasencia, N. Sendrier, J.P. Tillich);
- CT-RSA 2018: April 16-20, 2018, San Francisco, USA (M. Naya-Plasencia);
- Eurocrypt 2018: April 29- May 3, 2018, Tel Aviv, Israel (M. Naya-Plasencia);
- WAIFI 2018: June 14-16, 2018, Bergen, Norway, (L. Perrin)
- SAC 2018: August 13-14, 2018, Calgary, Canada, (G. Leurent);
- Crypto 2018: August 17-19, 2018, Santa Barbara, USA, (M. Naya-Plasencia);
- QCrypt 2018: August 27-31, 2018, Shanghai, China, (A. Leverrier);
- TQC 2018: July 16-18, 2018, Sydney, Australia, (A. Leverrier);
- QTech 2018: September 5-7, 2018, Paris, France, (A. Leverrier);
- SCN 2018: September 5-7, 2018, Amalfi, Italy, (G. Leurent);
- AQIS 2018: September 8-12, 2018, Nagoya, Japan, (A. Leverrier);
- SETA 2018: October 1-6, 2018, Hong-Kong, China, (P. Charpin);
- Asiacrypt 2018: December 02-06, 2018, Brisbane, Australia, (G. Leurent);
- CT-RSA 2019: March 4-8, 2019, San Francisco, USA, (L. Perrin)
- FSE 2019: March 25-28, 2019, Paris, France (C. Boura, A. Canteaut, G. Leurent, M. Naya-Plasencia)
- WCC 2019: March 31 - April 5, 2019, St Jacut-de-la-Mer, France, (A. Canteaut chair, P. Charpin, N. Sendrier, J.P. Tillich);
- PQCrypto 2019: May 8-10, 2019, Chongqing, China, (J.P. Tillich);
- CBC 2019: May 18-19, Darmstadt, Germany, (J.P. Tillich);
- Eurocrypt 2019: May 19-23, 2019, Darmstadt, Germany (C. Boura)
- ISIT 2019: July 7-12, 2019, Paris, France, (J.P. Tillich);
- CHES 2019: August 25-28, 2019, Atlanta, USA, (G. Leurent);
- Eurocrypt 2020: Zagreb, Croatia (A. Canteaut, PC co-chair).

### *9.1.3. Journal*

*9.1.3.1. Member of the Editorial Boards*

- *Designs, Codes and Cryptography*, associate editor: P. Charpin.
- *Finite Fields and Applications*, associate editor: A. Canteaut, P. Charpin.
- *Applicable Algebra in Engineering, Communication and Computing*, associate editor: A. Canteaut.
- *IACR Transactions on Symmetric Cryptology*, associate editors: C. Boura, A. Canteaut, G. Leurent, M. Naya-Plasencia.
- *IACR Transactions on Cryptographic Hardware and Embedded Systems*, associate editors: G. Leurent.
- *Advances in Mathematics of Communications*, associate editors: N. Sendrier and J.P. Tillich

*9.1.3.2. Reviewer - Reviewing Activities*

- Remote Referee - step 2- ERC-2018-CoG (A. Canteaut)
- Remote Referee - step 2- ERC-2018-STG (M. Naya-Plasencia)

### *9.1.4. Invited Talks*

- A. Canteaut, *Desperately Seeking Sboxes*, Eurocrypt 2018, Tel Aviv, Israel, April 29 - May 3 2018.
- M. Naya-Plasencia, *New Results on Quantum Symmetric Cryptanalysis*, QUANTALGO Quantum Algorithms and Applications Workshop, 2018, Paris, France, September 25 - 28, 2018.
- M. Naya-Plasencia, *New Results on Quantum Symmetric Cryptanalysis*, CrossFYRE Workshop, 2018, Surrey, UK, September 13 - 14, 2018.
- M. Naya-Plasencia, *New Results on Quantum Symmetric Cryptanalysis*, Journées Nationales 2018 du GDR Informatique Mathématique, Apr 2018, Palaiseau, France
- J.P. Tillich *Schémas cryptographiques à clé publique à base de codes correcteurs proposés à la compétition du NIST*, Journées Nationales 2018 du Pré-GDR Sécurité Informatique, June 1, 2018.

The members of the project-team have also been invited to give talks to some workshops or international seminars, including:

- C. Boura, A. Canteaut, J. Jean and V. Suder, *On Sboxes sharing the same DDT*, Dagstuhl Seminar 18021 Symmetric Cryptography, Jan 2018, Dagstuhl, Germany
- A. Canteaut *L'insoutenable légèreté du chiffrement*, Journées Scientifiques Inria 2018, June 2018, Bordeaux, France
- A. Canteaut and L. Perrin *On CCZ-Equivalence, Extended-Affine Equivalence and Function Twisting*, BFA 2018 - 3rd International Workshop on Boolean Functions and their Applications, Jun 2018, Loen, Norway
- A. Chailloux, *Relativistic commitment and zero-knowledge proofs*, 17th Bellairs Crypto-Workshop 2018, Mar 2018, Holetown, Barbados.
- T. Fuhr, M. Naya-Plasencia and Y. Rotella, *New Results on Modified Versions of Ketje Jr*, Dagstuhl Seminar 18021 Symmetric Cryptography, Jan 2018, Dagstuhl, Germany
- G. Leurent, *MDS Matrices with Lightweight Circuits*, The Challenges of Lightweight Cryptanalysis, April 2018, Tel Aviv, Israel.
- G. Leurent, *Security Issues with Small Block Sizes*, Lightweight Crypto Day, April 2018, Tel Aviv, Israel.
- G. Leurent *The Missing Difference Problem*, Flexible Symmetric Cryptography, March 2018, Leiden, Netherlands.
- M. Naya-Plasencia, *Quantum Safe Symmetric Cryptography*, Flexible Symmetric Cryptography Lorentz Center Workshop, 2018, Leiden, Netherlands, March 19 - 23, 2018.
- M. Naya-Plasencia, *Symmetric lightweight primitives: (Design and) Cryptanalysis*, Lightweight Crypto Day, April 2018, Tel Aviv, Israel.
- L. Perrin, *Generalized Feistel Networks with Optimal Diffusion*, Dagstuhl Seminar 18021 Symmetric Cryptography, Jan 2018, Dagstuhl, Germany
- L. Perrin, *S-Box Reverse-Engineering: Boolean Functions, American/Russian Standards, and Butterflies*, CECC 2018 - Central European Conference on Cryptology, Jun 2018, Smolenice, Slovakia.

### *9.1.5. Leadership within the Scientific Community*

- A. Canteaut serves as a chair of the steering committee of *Fast Software Encryption (FSE)*.
- A. Canteaut serves on the steering committee of the international competition CAESAR for authenticated encryption [0]
- N. Sendrier serves on the steering committee of *Post-quantum cryptography (PQCrypto)*.
- P. Charpin, N. Sendrier and JP Tillich serve on the steering committee of the WCC conference series.
- A. Leverrier serves on the steering committee of *DIM SIRTEQ* (réseau francilien pour les technologies quantiques).

[0] https://competitions.cr.yp.to/caesar.html

### 9.1.6. Research Administration

- A. Canteaut serves as Head of Science of the Inria Paris research center since September 2017.
- A. Canteaut serves on the *Inria Evaluation Committee* since September 2017.
- M. Naya-Plasencia and G. Leurent are members of *Inria Paris CSD Committee* (Comité de suivi doctoral).
- M. Naya-Plasencia is a member of *Inria Paris Scientific Hiring Committee* (Assignement of PhD, post-doctoral and delegation Inria fundings).
- M. Naya-Plasencia serves as head of the jury for PhD scholarships from EDITE.
- M. Naya-Plasencia serves on the *Comité des usagers du projet "rue Barrault"*.

### 9.1.7. Committees for the selection of professors, assistant professors and researchers

- Inria Paris Chargés de recherche: A. Canteaut (vice-chair)
- Inria Chargés de recherche (national selection): A. Canteaut
- ISTIC, Rennes, maître de conférence: M. Naya-Plasencia

## 9.2. Teaching - Supervision - Juries

### 9.2.1. Teaching

Master: A. Canteaut, *Error-correcting codes and applications to cryptology*, 12 hours, M2, University Paris-Diderot (MPRI), France;

Master: A. Chailloux, *Quantum Information*, 18 hours, M2, University Paris-Diderot (MPRI), France;

Master: A. Leverrier, *Quantum information and cryptography*, 18 hours, M2, University Paris-Diderot (MPRI), France;

Master: N. Sendrier, Information theory, 40 hours, M1, UVSQ, MINT, France;

Master: J.-P. Tillich, *Introduction to Information Theory*, 32 hours, M2, Ecole Polytechnique, France;

Corps des Mines: G. Leurent *Cryptographie symétrique*, 7 hours, Telecom ParisTech, France;

The members of the project-team were also invited to give courses at training schools for PhD students and young researchers:

- A. Canteaut, *Secure building-blocks against differential and linear attacks*, Training School on Symmetric Cryptography and Blockchain, Torremolinos, Spain, February 2018. 3 hours.
- A. Canteaut, *Exploiting algebraic properties of block ciphers*, Training School on Symmetric Cryptography and Blockchain, Torremolinos, Spain, February 2018. 1.5 hours.
- G. Leurent *How Not to Use a Blockcipher*, Training School on Symmetric Cryptography and Blockchain, Torremolinos, Spain, February 2018. 2.5 hours.
- A. Leverrier, *Security of continuous-variable quantum key distribution*, Secure Quantum Communications School, Baiona, Spain, May 2018.
- M. Naya-Plasencia, *Introduction to Symmetric Cryptography*, Summer School on real-world crypto and privacy, Sibenik, Croatia, June 2018.
- M. Naya-Plasencia, *Lightweight Cryptography*, Summer School on real-world crypto and privacy, Sibenik, Croatia, June 2018.

### *9.2.2. Supervision*

PhD: Sébastien Duval, *Constructions for lightweight cryptography*, Sorbonne Université, October 3, 2018.

PhDs: Yann Rotella, *Finite fields and symmetric cryptography*, Sorbonne Université, September 19, 2018.

PhD in progress: Rodolfo Canto Torres, *Analysis of generic decoding algorithms for the Hamming metric and study of cryptosystems based on the rank metric*, since September 2015, supervisor: N. Sendrier

PhD in progress: Xavier Bonnetain, *Cryptanalysis of symmetric primitives in the post-quantum world*, since September 2016, supervisor: M. Naya Plasencia

PhD in progress: Thomas Debris, *Quantum algorithms for decoding linear codes*, since September 2016, supervisor: J.-P. Tillich

PhD in progress: Antoine Grospellier, *LDPC codes: constructions and decoding*, since October 2016, supervisor: J.-P. Tillich

PhD in progress: Vivien Londe, *Study of quantum LDPC codes*, since September 2016, supervisors: G. Zémor and A. Leverrier

PhD in progress: Kevin Carrier, *Reconstruction of error-correcting codes*, since October 2016, supervisor: N. Sendrier

PhD in progress: Matthieu Lequesne, *Attaques par canaux cachés sur les cryptosystèmes à base de codes MDPC quasi-cycliques*, since September 2017, supervisor: N. Sendrier

PhD in progress: Ferdinand Sibleyras, *Security of modes of operation*, since October 2017, supervisor: G. Leurent and A. Canteaut

PhD in progress: Valentin Vasseur, *Etude du décodage des codes QC-MDPC*, since October 2017, supervisor: N. Sendrier

PhD in progress: Rémi Bricout, *Etude de scénarios non-locaux quantiques à l'aide d'outils de la théorie de l'information quantique*, since September 2017, supervisor: A. Chailloux and A. Leverrier

PhD in progress: Shouvik Ghorai, *Beyond-QKD continuous-variable quantum cryptographic protocols*, since October 2017, supervisors: E. Diamanti (UPMC), A. Leverrier

PhD in progress: Andrea Olivo, *Partir de contraintes relativistes pour faire de la cryptographie quantique*, since November 2017, supervisors: A. Chailloux and F. Grosshans (laboratoire Aimé Cotton).

PhD in progress: Daniel Coggia, *Cryptanalysis techniques for lightweight ciphers*, since September 2018, supervisors: A. Canteaut and C. Boura.

### *9.2.3. Juries*

- Alex Bredariol Grilo, *Quantum proofs, the Local Hamiltonian problem and applications*; Université Sorbonne Paris Cité, Paris, April 27, 2018, committee: A. Leverrier.
- Vincent Zucca, *Towards efficient arithmetic for Ring-LWE based homomorphic encryption*, Sorbonne Université, June 25, 2018, committee: A. Canteaut (chair);
- Yann Rotella, *Mathématiques discrètes appliquées à la cryptographie symétrique*, Sorbonne Université, September 19, 2018, committee: A. Canteaut (supervisor), M. Naya-Plasencia
- Dahmun Goudarzi, *Secure implementation of block ciphers against physical attacks*, PSL, September 21, 2018, committee: A. Canteaut
- Sébastien Duval, *Constructions pour la cryptographie à bas coût*, Sorbonne Université, October 3, committee: C. Boura, A. Canteaut (supervisor), G. Leurent (supervisor)

- Benjamin Lac, *Cryptographie légère intrinsèquement résistante aux attaques physiques pour l'Internet des objets*, Ecole des Mines de St-Etienne, October 18, 2018, committee: A. Canteaut
- Michele Minelli, *Chiffrement Totalement Homomorphe pour l'Apprentissage Automatique*, Université Paris Sciences et Lettres, October 26, 2018, committee: M. Naya-Plasencia (chair)
- Claire Delaplace, *Algorithmes d'algèbre linéaire pour la cryptographie*, Université de Rennes, November 21, 2018, committee: M. Naya-Plasencia.
- David Gérault, *Security Analysis of Contactless Communication Protocols*, Université Clermont Auvergne, November 27, 2018, committee: M. Naya-Plasencia (reviewer).
- Colin Chaigneau, *Cryptanalyse des Algorithmes de Chiffrement Symétrique*, Université de Versailles, November 28, 2018, committee: M. Naya-Plasencia (reviewer).
- Victor Cauchois, *Couches de Diffusion Lineaires à Partir de Matrices MDS*, Université de Rennes, December 13, 2018, committee: M. Naya-Plasencia.
- Eloi de Chérisey, *Towards a better formalisation of the side-channel threat*, Telecom Paris, December 18, 2018, committee: A. Canteaut (chair).

## 9.3. Popularization

### *9.3.1. Internal or external Inria responsibilities*

- **Association Animath**: M. Lequesne serves on the board of Animath.
- M. Lequesne is also member of the scientific committee of the French Tournament of Young Mathematicians: redaction of the problems for the competition, jury member (chair of a jury) ; member of the scientific committee of the International Tournament of Young Mathematicians: redaction of the problems for the competition, jury member (chair of a jury) ; Member of the scientific committee of the Correspondances des Jeunes Mathématicien.ne.s: redaction of the problems for the competition.

### *9.3.2. Articles and contents*

- A.Chailloux, *L'algorithme de Shor*, Interstices, Inria, March 2018.
- G. Leurent and M. Naya-Plasencia, *La fragilité inattendue du chiffrement symétrique*, "La Recherche", November 2018.
- JP Tillich, *Les codes correcteurs*, "La Recherche", November 2018, p. 45-46.
- A. Canteaut, *La meilleure garantie de sécurité est l'épreuve du temps*, interview to the journal "La Recherche", November 2018.
- M. Naya-Plasencia, *Symmetric Cryptanalysis: The Foundation of Trust*, Lorentz Center Highlights, 2018, Leiden, Netherlands, Mars 20, 2018.

### *9.3.3. Education*

- **Alkindi cipher challenge:** Several members of the project-team are involved in the cipher challenge for high-school students "concours Alkindi" http://www.concours-alkindi.fr/. Mathieu Lequesne serves as a co-organizer of the challenge, preparing the three rounds and the final. Together with C. Boura and A. Canteaut, he was also involved in the redaction of the exercises, and in videos for Inria channel on different aspects of cryptography and how to solve problems from the Alkindi challenge: https://www.youtube.com/watch?v=Y-VQBzwEaqQ&t=17s, https://www.youtube.com/watch?v=Mv415zfUFNs&t=3s and https://www.youtube.com/watch?v=8ohEeTPKBwA&t=21s. The best teams from Académie de Paris have been visiting the SECRET project-team in June 2018 https://www.youtube.com/watch?v=EVLHEOWAORc.
- Organization of the event "Rendez-vous des Jeunes Mathématiciennes et Informaticiennes" at Inria Paris (October 22-23) by M. Lequesne, a 2-days camp for 20 high-school girls interested in mathematics and computer science.

- Organization of the International Tournament of Young Mathematicians in Paris, a one-week competition (July 5-12) for 120 high-school students. M. Lequesne served as vice-president of the local organizing committee.

### *9.3.4. Interventions*

- A. Canteaut gave a talk to high-school students at Palais de la Découverte, during the "Semaine des maths" (March 2018) [61];
- A. Canteaut gave the talk during the closing ceremony of "Olympiades nationales de mathématiques" (June 2018) [62];
- A. Canteaut gave a presentation on research in computer science to 10-year children in a school in Paris (Jan. 2018);
- M. Lequesne gave a presentation on code-based cryptography to high-school interns (stagiaires de 3e) (Dec. 2018).

# 10. Bibliography

## Major publications by the team in recent years

[1] C. BEIERLE, A. CANTEAUT, G. LEANDER, Y. ROTELLA.*Proving Resistance Against Invariant Attacks: How to Choose the Round Constants*, in "Crypto 2017 - Advances in Cryptology", Santa Barbara, United States, J. KATZ, H. SHACHAM (editors), LNCS - Lecture Notes in Computer Science, Springer, August 2017, vol. 10402, p. 647–678 [*DOI :* 10.1007/978-3-319-63715-0_22], https://hal.inria.fr/hal-01631130

[2] K. BHARGAVAN, G. LEURENT.*On the Practical (In-)Security of 64-bit Block Ciphers*, in "ACM CCS 2016 - 23rd ACM Conference on Computer and Communications Security", Vienna, Austria, ACM, October 2016 [*DOI :* 10.1145/2976749.2978423], https://hal.inria.fr/hal-01404208

[3] A. CANTEAUT, J. ROUÉ.*On the behaviors of affine equivalent Sboxes regarding differential and linear attacks*, in "Advances in Cryptology - Eurocrypt 2015", Sofia, Bulgaria, Lecture Notes in Computer Science, Springer, April 2015, https://hal.inria.fr/hal-01104051

[4] A. CHAILLOUX, M. NAYA-PLASENCIA, A. SCHROTTENLOHER.*An Efficient Quantum Collision Search Algorithm and Implications on Symmetric Cryptography*, in "Asiacrypt 2017 - Advances in Cryptology", Hong Kong, China, T. TAKAGI, T. PEYRIN (editors), LNCS - Lecture Notes in Computer Science, Springer, December 2017, vol. 10625, p. 211–240 [*DOI :* 10.1007/978-3-319-70697-9_8], https://hal.inria.fr/hal-01651007

[5] K. CHAKRABORTY, A. CHAILLOUX, A. LEVERRIER.*Arbitrarily long relativistic bit commitment* , in "Physical Review Letters", 2015 [*DOI :* 10.1103/PHYSREVLETT.115.250501], https://hal.inria.fr/hal-01237241

[6] P. CHARPIN, G. M. KYUREGHYAN, V. SUDER.*Sparse Permutations with Low Differential Uniformity*, in "Finite Fields and Their Applications", March 2014, vol. 28, p. 214-243 [*DOI :* 10.1016/J.FFA.2014.02.003], https://hal.archives-ouvertes.fr/hal-01068860

[7] N. COURTOIS, M. FINIASZ, N. SENDRIER.*How to achieve a McEliece-based Digital Signature Scheme*, in "Advances in Cryptology - Asiacrypt 2001", LNCS, Springer-Verlag, 2001, n$^{o}$ 2248, p. 157–174

[8] A. COUVREUR, A. OTMANI, J.-P. TILLICH.*Polynomial Time Attack on Wild McEliece Over Quadratic Extensions*, in "IEEE Transactions on Information Theory", January 2017, vol. 63, n° 1, p. 404–427 [*DOI :* 10.1109/TIT.2016.2574841], https://hal.inria.fr/hal-01661935

[9] M. KAPLAN, G. LEURENT, A. LEVERRIER, M. NAYA-PLASENCIA.*Breaking Symmetric Cryptosystems Using Quantum Period Finding*, in "Crypto 2016 - 36th Annual International Cryptology Conference", Santa Barbara, United States, M. ROBSHAW, J. KATZ (editors), LNCS - Lecture Notes in Computer Science, Springer, August 2016, vol. 9815, p. 207 - 237 [*DOI :* 10.1007/978-3-662-53008-5_8], https://hal.inria.fr/hal-01404196

[10] R. MISOCZKI, J.-P. TILLICH, N. SENDRIER, P. S. L. M. BARRETO.*MDPC-McEliece: New McEliece Variants from Moderate Density Parity-Check Codes*, in "IEEE International Symposium on Information Theory - ISIT 2013", Istanbul, Turkey, July 2013, p. 2069-2073, https://hal.inria.fr/hal-00870929

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] S. DUVAL.*Constructions for Lightweight Cryptography*, Sorbonne Université , UPMC, October 2018, https://hal.inria.fr/tel-01900290

[12] Y. ROTELLA.*Discrete Mathematics for symmetric cryptography*, Sorbonne Université, September 2018, https://hal.inria.fr/tel-01944827

### Articles in International Peer-Reviewed Journal

[13] C. BEIERLE, A. CANTEAUT, G. LEANDER.*Nonlinear Approximations in Cryptanalysis Revisited*, in "IACR Transactions on Symmetric Cryptology", December 2018, vol. 2018, n° 4, p. 80-101 [*DOI :* 10.13154/TOSC.V2018.I4.80-101], https://hal.inria.fr/hal-01944995

[14] C. BOURA, A. CANTEAUT.*On the Boomerang Uniformity of Cryptographic Sboxes*, in "IACR Transactions on Symmetric Cryptology", September 2018, vol. 2018, n° 3, p. 290-310 [*DOI :* 10.13154/TOSC.V2018.I3.290-310], https://hal.inria.fr/hal-01944598

[15] C. BOURA, A. CANTEAUT, J. JEAN, V. SUDER.*Two Notions of Differential Equivalence on Sboxes*, in "Designs, Codes and Cryptography", 2018 [*DOI :* 10.1007/S10623-018-0496-Z], https://hal.inria.fr/hal-01944565

[16] C. BOURA, V. LALLEMAND, V. SUDER, M. NAYA-PLASENCIA.*Making the Impossible Possible*, in "Journal of Cryptology", January 2018, vol. 31, n° 1, p. 101-133 [*DOI :* 10.1007/S00145-016-9251-7], https://hal.inria.fr/hal-01953916

[17] A. CANTEAUT, S. CARPOV, C. FONTAINE, T. LEPOINT, M. NAYA-PLASENCIA, P. PAILLIER, R. SIRDEY.*Stream Ciphers: A Practical Solution for Efficient Homomorphic-Ciphertext Compression*, in "Journal of Cryptology", July 2018, vol. 31, n° 3, p. 885-916 [*DOI :* 10.1007/S00145-017-9273-9], https://hal.inria.fr/hal-01650012

[18] A. CANTEAUT, L. PERRIN.*On CCZ-Equivalence, Extended-Affine Equivalence, and Function Twisting*, in "Finite Fields and Their Applications", March 2019, vol. 56, p. 209-246 [*DOI :* 10.1016/J.FFA.2018.11.008], https://hal.inria.fr/hal-01953353

[19] P. CHARPIN, J. PENG.*New links between nonlinearity and differential uniformity*, in "Finite Fields and Their Applications", March 2019, vol. 56, p. 188-208 [*DOI :* 10.1016/J.FFA.2018.12.001], https://hal.inria.fr/hal-01907499

[20] S. DUVAL, G. LEURENT.*MDS Matrices with Lightweight Circuits*, in "IACR Transactions on Symmetric Cryptology", June 2018 [*DOI :* 10.13154/TOSC.V2018.I2.48-78], https://hal.inria.fr/hal-01944495

[21] T. FUHR, M. NAYA-PLASENCIA, Y. ROTELLA.*State-Recovery Attacks on modified Ketje Jr*, in "IACR Transactions on Symmetric Cryptology", March 2018, vol. 2018, n$^{O}$ 1, p. 29-56 [*DOI :* 10.13154/TOSC.V2018.I1.29-56], https://hal.inria.fr/hal-01944785

[22] S. GHORAI, E. DIAMANTI, A. LEVERRIER.*Composable security of two-way continuous-variable quantum key distribution without active symmetrization*, in "Physical Review A", 2019, https://arxiv.org/abs/1806.11356 [*DOI :* 10.1103/PHYSREVA.99.012311], https://hal.inria.fr/hal-01951932

[23] A. LEVERRIER.$SU(p,q)$ *coherent states and a Gaussian de Finetti theorem*, in "Journal of Mathematical Physics", 2018, vol. 59, 042202, https://arxiv.org/abs/1612.05080 [*DOI :* 10.1063/1.5007334], https://hal.inria.fr/hal-01652084

[24] A. OLIVO, F. GROSSHANS.*Ancilla-assisted linear optical Bell measurements and their optimality*, in "Physical Review A", October 2018, vol. 98, n$^{O}$ 4, 042323 [*DOI :* 10.1103/PHYSREVA.98.042323], https://hal.inria.fr/hal-01951361

## Invited Conferences

[25] C. BOURA, A. CANTEAUT, J. JEAN, V. SUDER.*On Sboxes sharing the same DDT*, in "Dagstuhl Seminar 18021 Symmetric Cryptography", Dagstuhl, Germany, January 2018 [*DOI :* 10.4230/DAGREP.8.1.1], https://hal.inria.fr/hal-01955256

[26] A. CANTEAUT, C. BEIERLE, G. LEANDER.*On nonlinear approximations and the linear hull effect*, in "ASK 2018 - 8th Asian Workshop on Symmetric Key Cryptography", Kolkata, India, November 2018, https://hal.inria.fr/hal-01955286

[27] A. CANTEAUT.*Desperately Seeking Sboxes*, in "Eurocrypt 2018", Tel Aviv, Israel, April 2018, https://hal.inria.fr/hal-01944401

[28] A. CANTEAUT.*L'insoutenable légèreté du chiffrement*, in "Journées Scientifiques Inria 2018", Bordeaux, France, June 2018, https://hal.inria.fr/hal-01955337

[29] A. CANTEAUT, L. PERRIN.*On CCZ-Equivalence, Extended-Affine Equivalence and Function Twisting*, in "BFA 2018 - 3rd International Workshop on Boolean Functions and their Applications", Loen, Norway, June 2018, https://hal.inria.fr/hal-01953349

[30] A. CHAILLOUX.*Relativistic commitment and zero-knowledge proofs*, in "Seventeenth Bellairs Crypto-Workshop 2018", Holetown, Barbados, March 2018, https://hal.inria.fr/hal-01950643

[31] G. LEURENT.*MDS Matrices with Lightweight Circuits*, in "The Challenges of Lightweight Cryptanalysis", Tel Aviv, Israel, April 2018, https://hal.inria.fr/hal-01953383

[32] G. LEURENT.*Security Issues with Small Block Sizes*, in "Lightweight Crypto Day 2018", Tel Aviv, Israel, April 2018, https://hal.inria.fr/hal-01966550

[33] A. LEVERRIER.*Introduction to quantum computing*, in "Lecture series on Quantum Engineering at University Paris-Saclay", Palaiseau, France, May 2018, https://hal.inria.fr/hal-01955373

[34] M. NAYA-PLASENCIA, T. FUHR, Y. ROTELLA.*New Results on Modified Versions of Ketje Jr*, in "Dagstuhl Seminar 18021 Symmetric Cryptography", Dagstuhl, Germany, January 2018 [*DOI :* 10.4230/DAGREP.8.1.1], https://hal.archives-ouvertes.fr/hal-01953975

[35] M. NAYA-PLASENCIA.*New Results on Quantum Symmetric Cryptanalysis*, in "Journées Nationales 2018 du GDR Informatique Mathématique", Palaiseau, France, April 2018, https://hal.inria.fr/hal-01954618

[36] M. NAYA-PLASENCIA.*New results on symmetric quantum cryptanalysis (Keynote speaker)*, in "QUANTALGO Quantum Algorithms and Applications", Paris, France, September 2018, https://hal.inria.fr/hal-01953994

[37] M. NAYA-PLASENCIA.*New results on symmetric quantum cryptanalysis*, in "Crossfyre 2018 - 8th international workshop on cryptography, robustness, and provably secure schemes for female young researchers", Surrey, United Kingdom, September 2018, Keynote speaker at Crossfyre 2018, https://hal.inria.fr/hal-01953997

[38] M. NAYA-PLASENCIA.*Symmetric lightweight primitives: (Design and) Cryptanalysis*, in "Lightweight Crypto Day 2018", Tel Aviv, Israel, April 2018, https://hal.inria.fr/hal-01953947

[39] L. PERRIN.*Generalized Feistel Networks with Optimal Diffusion*, in "Dagstuhl Seminar 18021 Symmetric Cryptography", Dagstuhl, Germany, January 2018 [*DOI :* 10.4230/DAGREP.8.1.1], https://hal.inria.fr/hal-01953351

[40] L. PERRIN.*S-Box Reverse-Engineering: Boolean Functions, American/Russian Standards, and Butterflies*, in "CECC 2018 - Central European Conference on Cryptology", Smolenice, Slovakia, June 2018, p. 1-99, https://hal.inria.fr/hal-01953348

### International Conferences with Proceedings

[41] N. ARAGON, P. GABORIT, A. HAUTEVILLE, J.-P. TILLICH.*A New Algorithm for Solving the Rank Syndrome Decoding Problem*, in "ISIT 2018 - IEEE International Symposium on Information Theory", Vail, United States, June 2018, p. 2421-2425 [*DOI :* 10.1109/ISIT.2018.8437464], https://hal.inria.fr/hal-01957179

[42] T. ASHUR, M. EICHLSEDER, M. M. LAURIDSEN, G. LEURENT, B. MINAUD, Y. ROTELLA, Y. SASAKI, B. VIGUIER.*Cryptanalysis of MORUS*, in "ASIACRYPT 2018 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, LNCS - Lecture Notes in Computer Science, Springer, December 2018, vol. 11273, p. 35-64 [*DOI :* 10.1007/978-3-030-03329-3_2], https://hal.inria.fr/hal-01944776

[43] X. BONNETAIN, M. NAYA-PLASENCIA.*Hidden Shift Quantum Cryptanalysis and Implications*, in "ASIACRYPT 2018 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, LNCS - Lecture Notes in Computer Science, Springer, December 2018, vol. 11272, p. 560-592 [*DOI :* 10.1007/978-3-030-03326-2_19], https://hal.inria.fr/hal-01953914

[44] G. COUTEAU, A. DUPIN, P. MÉAUX, M. ROSSI, Y. ROTELLA.*On the Concrete Security of Goldreich's Pseudorandom Generator*, in "ASIACRYPT 2018 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, LNCS - Lecture Notes in Computer Science, Springer, December 2018, vol. 11273, p. 96-124 [*DOI :* 10.1007/978-3-030-03329-3_4], https://hal.inria.fr/hal-01944772

[45] T. DEBRIS-ALAZARD, J.-P. TILLICH.*Two attacks on rank metric code-based schemes: RankSign and an IBE scheme*, in "ASIACRYPT 2018 - 24th International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, LNCS - Lecture Notes in Computer Science, Springer, December 2018, vol. 11272, p. 62-92 [*DOI :* 10.1007/978-3-030-03326-2_3], https://hal.inria.fr/hal-01957207

[46] E. EATON, M. LEQUESNE, A. PARENT, N. SENDRIER.*QC-MDPC: A Timing Attack and a CCA2 KEM*, in "PQCrypto 2018 - Ninth International Conference on Post-Quantum Cryptography", Fort Lauderdale, United States, LNCS - Lecture Notes in Computer Science, Springer, April 2018, vol. 10786 [*DOI :* 10.1007/978-3-319-79063-3_3], https://hal.inria.fr/hal-01949590

[47] O. FAWZI, A. GROSPELLIER, A. LEVERRIER.*Constant overhead quantum fault-tolerance with quantum expander codes*, in "FOCS 2018 - 59th Annual IEEE Symposium on Foundations of Computer Science", Paris, France, October 2018, p. 743-754, https://arxiv.org/abs/1808.03821 [*DOI :* 10.1109/FOCS.2018.00076], https://hal.archives-ouvertes.fr/hal-01895430

[48] O. FAWZI, A. GROSPELLIER, A. LEVERRIER.*Efficient decoding of random errors for quantum expander codes*, in "STOC 2018 - 50th Annual ACM Symposium on the Theory of Computing", Los Angeles, United States, June 2018, p. 521-534, https://arxiv.org/abs/1711.08351 [*DOI :* 10.1145/3188745.3188886], https://hal.archives-ouvertes.fr/hal-01895427

[49] L. GRASSI, M. NAYA-PLASENCIA, A. SCHROTTENLOHER.*Quantum Algorithms for the k-xor Problem*, in "ASIACRYPT 2018 - 24th Annual International Conference on the Theory and Application of Cryptology and Information Security", Brisbane, Australia, LNCS - Lecture Notes in Computer Science, Springer, December 2018, vol. 11272, p. 527-559 [*DOI :* 10.1007/978-3-030-03326-2_18], https://hal.inria.fr/hal-01896036

[50] B. LAC, A. CANTEAUT, J. J.-A. FOURNIER, R. SIRDEY.*Thwarting Fault Attacks against Lightweight Cryptography using SIMD Instructions*, in "ISCAS 2018 - IEEE International Symposium on Circuits and Systems", Florence, Italy, May 2018, p. 1-5 [*DOI :* 10.1109/ISCAS.2018.8351693], https://hal-cea.archives-ouvertes.fr/cea-01746138

[51] M. LEQUESNE, J.-P. TILLICH.*Attack on the Edon-K Key Encapsulation Mechanism*, in "ISIT 2018 - IEEE International Symposium on Information Theory", Vail, United States, June 2018, p. 981-985 [*DOI :* 10.1109/ISIT.2018.8437498], https://hal.inria.fr/hal-01949569

[52] G. LEURENT, M. NANDI, F. SIBLEYRAS.*Generic Attacks Against Beyond-Birthday-Bound MACs*, in "Crypto 2018 - 38th International Cryptology Conference", Santa Barbara, United States, LNCS - Lecture Notes in Computer Science, Springer, August 2018, vol. 10991, p. 306-336 [*DOI :* 10.1007/978-3-319-96884-1_11], https://hal.inria.fr/hal-01944318

[53] G. LEURENT, F. SIBLEYRAS.*The Missing Difference Problem, and Its Applications to Counter Mode Encryption*, in "Eurocrypt 2018 - 37th Annual International Conference on the Theory and Applications of

Cryptographic Techniques", Tel Aviv, Israel, LNCS - Lecture Notes in Computer Science, April 2018, vol. 10821, p. 745-770 [*DOI :* 10.1007/978-3-319-78375-8_24], https://hal.inria.fr/hal-01944288

[54] J.-P. TILLICH.*The decoding failure probability of MDPC codes*, in "ISIT 2018 - IEEE International Symposium on Information Theory", Vail, United States, June 2018, p. 941-945 [*DOI :* 10.1109/ISIT.2018.8437843], https://hal.inria.fr/hal-01957037

### Conferences without Proceedings

[55] P. CHARPIN, J. PENG.*New links between nonlinearity and differential uniformity*, in "Sequences and Their Applications (SETA) 2018", Hong-Kong, China, October 2018, https://hal.inria.fr/hal-01836184

[56] O. FAWZI, A. GROSPELLIER, A. LEVERRIER.*Efficient decoding of random errors for quantum expander codes*, in "QIP 2018 - 21th Annual Conference on Quantum Information Processing", Delft, Netherlands, QuTech, January 2018, p. 1-31, https://arxiv.org/abs/1711.08351 - 31 pages, https://hal.archives-ouvertes.fr/hal-01654670

[57] G. LEURENT.*The Missing Difference Problem: And its Applications to Counter Mode Encryption*, in "Flexible Symmetric Cryptography", Leiden, Netherlands, March 2018, https://hal.inria.fr/hal-01953390

[58] A. OLIVO, F. GROSSHANS.*Optimality of linear optical Bell measurements. How much can ancillae help?*, in "GDR IQFA 9th Colloquium", Montpellier, France, November 2018, https://hal.inria.fr/hal-01951749

[59] A. OLIVO, F. GROSSHANS.*Optimality of linear optical Bell measurements. How much can ancillae help?*, in "ICIQP 2018 - International Conference on Integrated Quantum Photonics", Paris, France, October 2018, https://hal.inria.fr/hal-01951728

[60] A. OLIVO, F. GROSSHANS.*Optimality of linear optical Bell measurements. How much can ancillae help?*, in "Q-Turn: changing paradigms in quantum science", Florianopolis, Brazil, November 2018, https://hal.inria.fr/hal-01951753

### Scientific Popularization

[61] A. CANTEAUT.*Chut ! On nous écoute*, in "Semaine des Maths 2018", Paris, France, March 2018, https://hal.inria.fr/hal-01955267

[62] A. CANTEAUT.*Chut ! On nous écoute*, in "Conférence de clôture des Olympiades Nationales de Mathématiques 2018", Paris, France, June 2018, https://hal.inria.fr/hal-01955273

[63] A. CHAILLOUX.*L'algorithme quantique de Shor*, in "Interstices", March 2018, https://hal.inria.fr/hal-01827601

[64] G. LEURENT, M. NAYA-PLASENCIA.*La fragilité inattendue du chiffrement symétrique*, in "La Recherche : l'actualité des sciences", November 2018, vol. Novembre 2018, https://hal.inria.fr/hal-01953448

[65] L. PERRIN.*Building Light but not Weak Protections for the IoT*, in "PhD Graduation Ceremony of the University of Luxembourg (2018)", Belval, Luxembourg, December 2018, https://hal.inria.fr/hal-01959751

## Other Publications

[66] X. BONNETAIN, M. NAYA-PLASENCIA, A. SCHROTTENLOHER.*On Quantum Slide Attacks*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01946399

[67] X. BONNETAIN, M. NAYA-PLASENCIA, A. SCHROTTENLOHER.*Quantum Cryptanalysis of AES*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01955534

[68] X. BONNETAIN, A. SCHROTTENLOHER.*Quantum Security Analysis of CSIDH and Ordinary Isogeny-based Schemes*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01896046

[69] X. BONNETAIN, A. SCHROTTENLOHER.*Submerging CSIDH*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01961633

[70] A. CANTEAUT.*Exploiting algebraic properties of block ciphers*, February 2018, COST Training School on Symmetric Cryptography and Blockchain, Torremolinos, Spain, https://hal.inria.fr/hal-01955320

[71] A. CANTEAUT.*Secure building-blocks against differential and linear attacks*, February 2018, COST Training School on Symmetric Cryptography and Blockchain, Torremolinos, Spain, https://hal.inria.fr/hal-01955315

[72] A. CANTEAUT, L. PERRIN.*On CCZ-Equivalence, Extended-Affine Equivalence and Function Twisting*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01959749

[73] K. CARRIER, J.-P. TILLICH.*Near collisions search and generic decoding*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01959614

[74] A. CHAILLOUX.*A note on the quantum query complexity of permutation symmetric functions*, December 2018, https://arxiv.org/abs/1810.01790 - 8 pages [*DOI :* 10.01790], https://hal.inria.fr/hal-01950650

[75] A. CHAILLOUX.*DEREC - Développement de la cryptographie relativiste*, October 2018, WISG 2018 - 12ème Workshop Interdisciplinaire sur la Sécurité Globale, Poster, https://hal.inria.fr/hal-01950649

[76] P. CHARPIN, J. PENG.*Differential uniformity and the associated codes of cryptographic functions*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01908336

[77] D. COGGIA.*On subspace trails cryptanalysis*, Université Paris Diderot (Paris 7), September 2018, https://hal.inria.fr/hal-01955305

[78] D. COGGIA.*On subspace trails cryptanalysis*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01960306

[79] A. COUVREUR, M. LEQUESNE, J.-P. TILLICH.*Recovering short secret keys of RLCE encryption scheme in polynomial time*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01959617

[80] A. COUVREUR, M. LEQUESNE, J.-P. TILLICH.*Recovering short secret keys of RLCE in polynomial time*, May 2018, https://arxiv.org/abs/1805.11489 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01803440

[81] T. DEBRIS-ALAZARD, N. SENDRIER, J.-P. TILLICH.*Wave: A New Code-Based Signature Scheme*, December 2018, preprint IACR disponible sur https://eprint.iacr.org/2018/996/20181022:154324, https://hal.inria.fr/hal-01958175

[82] T. DEBRIS-ALAZARD, J.-P. TILLICH.*Deux attaques contre des schémas se fondant sur les codes en métrique rang : Ranksign et un chiffrement basé sur l'identité*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01959613

[83] A. GROSPELLIER, A. KRISHNA.*Numerical estimate of the threshold for quantum expander codes*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.archives-ouvertes.fr/hal-01955453

[84] A. GROSPELLIER, A. KRISHNA.*Numerical study of hypergraph product codes*, October 2018, https://arxiv.org/abs/1810.03681 - 10 pages, 2 figures [*DOI :* 10.03681], https://hal.archives-ouvertes.fr/hal-01895436

[85] M. LEQUESNE, J.-P. TILLICH.*Attack on the EDON-K Key Encapsulation Mechanism*, November 2018, https://arxiv.org/abs/1802.06157 - Submitted to ISIT 2018, https://hal.sorbonne-universite.fr/hal-01925323

[86] G. LEURENT.*How Not to Use a Blockcipher*, February 2018, COST Training School on Symmetric Cryptography and Blockchain, Torremolinos, Spain, https://hal.inria.fr/hal-01953398

[87] G. LEURENT, F. SIBLEYRAS.*The Missing Difference Problem, and its Applications to Counter Mode Encryption*, October 2018, JC2 2018 - Journées Codage et Cryptographie, https://hal.inria.fr/hal-01961739

[88] A. LEVERRIER.*Security of continuous-variable quantum key distribution*, May 2018, Secure Quantum Communications School, Baiona, Spain, https://hal.inria.fr/hal-01955365

[89] F. MENDEL, M. NAYA-PLASENCIA.*Preface*, March 2018, vol. 2018, n$^{o}$ 1, p. 1 - 4, IACR Transactions on Symmetric Cryptology (ToSC) [*DOI :* 10.13154/TOSC.V2018.I1.1-4], https://hal.inria.fr/hal-01953923

[90] M. NAYA-PLASENCIA.*Introduction to Symmetric Cryptography*, June 2018, Summer School on real-world crypto and privacy, https://hal.inria.fr/hal-01953897

[91] M. NAYA-PLASENCIA.*Lightweight Cryptography*, June 2018, Summer School on real-world crypto and privacy, https://hal.inria.fr/hal-01953789

[92] M. NAYA-PLASENCIA.*New results on symmetric quantum cryptanalysis*, March 2018, Keynote speaker at Flexible symmetric cryptography -Lorentz Center, https://hal.inria.fr/hal-01954599

[93] M. NAYA-PLASENCIA.*New results on symmetric quantum cryptanalysis*, March 2018, Seminaire CCA, https://hal.inria.fr/hal-01954616

[94] M. NAYA-PLASENCIA.*Symmetric Cryptanalysis: the Foundation of Trust*, March 2018, Lorentz Center Highlights, https://hal.inria.fr/hal-01954612

[95] A. QUEROL CRUZ.*Conditional Differential Cryptanalysis of the Post-Quantum ARX Symmetric Primitive Salsa20*, Univeristé Denis Diderot Paris 7, September 2018, https://hal.inria.fr/hal-01893824

# Project-Team SERENA

## Simulation for the Environment: Reliable and Efficient Numerical Algorithms

IN COLLABORATION WITH: Centre d'Enseignement et de Recherche en Mathématiques et Calcul Scientifique (CERMICS)

IN PARTNERSHIP WITH:
**Ecole des Ponts ParisTech**

RESEARCH CENTER
**Paris**

THEME
**Earth, Environmental and Energy Sciences**

# Table of contents

# Project-Team SERENA

*Creation of the Team: 2015 June 01, updated into Project-Team: 2017 April 01*

**Keywords:**

**Computer Science and Digital Science:**

A2.1.3. - Object-oriented programming
A2.1.4. - Functional programming
A2.4.3. - Proofs
A6.1.1. - Continuous Modeling (PDE, ODE)
A6.1.4. - Multiscale modeling
A6.1.5. - Multiphysics modeling
A6.2.1. - Numerical analysis of PDE and ODE
A6.2.5. - Numerical Linear Algebra
A6.2.8. - Computational geometry and meshes
A6.3.1. - Inverse problems
A6.3.4. - Model reduction
A6.3.5. - Uncertainty Quantification

**Other Research Topics and Application Domains:**

B3.1. - Sustainable development
B3.3.1. - Earth and subsoil
B3.4.2. - Industrial risks and waste
B3.4.3. - Pollution
B4.1. - Fossile energy production (oil, gas)
B4.2.1. - Fission
B5.5. - Materials

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Martin Vohralík [Team leader, Inria, Senior Researcher, HDR]
François Clément [Inria, Researcher]
Alexandre Ern [Ecole Nationale des Ponts et Chaussées, Researcher, HDR]
Michel Kern [Inria, Researcher]
Géraldine Pichot [Inria, Researcher]
Pierre Weis [Inria, Senior Researcher]

**External Collaborators**

Hend Ben Ameur [IPEST and ENIT-Lamsin (Tunisia), Professor, HDR]
Guy Chavent [Univ Paris-Dauphine, Professor (retired), HDR]
Jérôme Jaffré [Inria, Senior Researcher (retired), HDR]
Caroline Japhet [Univ Paris-Nord, Associate Professor]
Antoine Lejay [Inria, Senior Researcher, HDR]
Lionel Lenôtre [Univ de Lorraine, IECL, Post-Doctoral Fellow]
Vincent Martin [Univ de technologie de Compiègne, Associate Professor]
Jean-Elizabeth Roberts [Inria, Senior Researcher (retired), HDR]

**Technical Staff**
Sébastien Furic [Inria]
Florent Hedin [Inria, from Dec 2018]
Simon Legrand [Inria]

**PhD Students**
Amina Benaceur [EDF]
Karol Cascavita [Univ Paris-Est]
Jad Dabaghi [Inria]
Patrik Daniel [Inria]
Frédéric Marazzato [CEA]
Riccardo Milani [EDF]
Ani Miraci [Inria]
Nicolas Pignet [EDF]

**Post-Doctoral Fellows**
Sarah Ali Hassan [Inria, from Apr 2018 until Aug 2018]
Matteo Cicuttin [Ecole Nationale des Ponts et Chaussées, until Apr 2018]
Théophile Chaumont-Frelet [Ecole Nationale des Ponts et Chaussées, from Feb to Sep 2018]
Guillaume Delay [Ecole Nationale des Ponts et Chaussées, from Sep 2018]
Kenan Kergrene Profit [Inria, from Dec 2018]
Seyed Mohammad Zakerzadeh [Inria]

**Visiting Scientists**
Carsten Carstensen [Humboldt University, Berlin, from Aug to Sep 2018, HDR]
Thirupathi Gudi [Indian Institute of Science, Bangalore, from Jan to Feb 2018, HDR]
Jean-Luc Guermond [Texas A&M University, from May 2018 until Jun 2018, HDR]
Christian Kreuzer [University College London, Jun 2018, HDR]
Iain Smears [University College London, Mar and Jun 2018]

**Administrative Assistants**
Virginie Collette [Inria, until Oct 2018]
Meriem Henni [Inria, from Apr 2018 until Aug 2018]
Derya Gök [Inria, from Nov 2018]

# 2. Overall Objectives

## 2.1. Overall Objectives

The project-team SERENA is concerned with **numerical methods** for **environmental problems**. The main topics are the conception and analysis of *models* based on *partial differential equations*, the study of their *precise and efficient numerical approximation*, and implementation issues with special concern for *reliability and correctness of programs*. We are in particular interested in *guaranteeing* the *quality* of the *overall simulation process*. SERENA has taken over the project-team POMDAPI2 which ended on May 31, 2015. It has been given an authorization to become a joint project-team between Inria and ENPC at the Committee of Projects, September 1st, 2016, and was created as project-team on April 10, 2017.

# 3. Research Program

## 3.1. Multiphysics coupling

Within our project, we start from the conception and analysis of *models* based on *partial differential equations* (PDEs). Already at the PDE level, we address the question of *coupling* of different models; examples are that of simultaneous fluid flow in a discrete network of two-dimensional *fractures* and in the surrounding three-dimensional porous medium, or that of interaction of a compressible flow with the surrounding elastic *deformable structure*. The key physical characteristics need to be captured, whereas existence, uniqueness, and continuous dependence on the data are minimal analytic requirements that we seek to satisfy. At the modeling stage, we also develop model-order reduction techniques, such as the use of reduced basis techniques or proper generalized decompositions, to tackle evolutive problems, in particular in the nonlinear case.

## 3.2. Structure-preserving discretizations and discrete element methods

We consequently design *numerical methods* for the devised model. Traditionally, we have worked in the context of finite element, finite volume, mixed finite element, and discontinuous Galerkin methods. Novel classes of schemes enable the use of general *polygonal* and *polyhedral meshes* with *nonmatching interfaces*, and we develop them in response to a high demand from our industrial partners (namely EDF, CEA, and IFP Energies Nouvelles). In the lowest-order case, our requirement is to derive *structure-preserving* methods, i.e., methods that mimic algebraically at the discrete level fundamental properties of the underlying PDEs, such as conservation principles and preservation of invariants. Here, the theoretical questions are closely linked to *differential geometry* and we apply them to the Navier–Stokes equations and to elasto-plasticity. In the higher-order case, we actively contribute to the development of hybrid high-order methods. We contribute to the numerical analysis in nonlinear cases (obstacle problem, Signorini conditions), we apply these methods to challenging problems from solid mechanics involving large deformations and plasticity, and we develop a comprehensive software implementing them. We believe that these methods belong to the future generation of numerical methods for industrial simulations; as a concrete example, the implementation of these methods in an industrial software of EDF has begun this year.

## 3.3. Domain decomposition and Newton–Krylov (multigrid) solvers

We next concentrate an intensive effort on the development and analysis of efficient solvers for the systems of nonlinear algebraic equations that result from the above discretizations. We have in the past developed *Newton–Krylov solvers* like the adaptive inexact Newton method, and we place a particular emphasis on *parallelization* achieved via the *domain decomposition* method. Here we traditionally specialize in *Robin transmission conditions*, where an optimized choice of the parameter has already shown speed-ups in orders of magnitude in terms of the number of domain decomposition iterations in model cases. We concentrate in the SERENA project on adaptation of these algorithms to the above novel discretization schemes, on the optimization of the free Robin parameter for challenging situations, and also on the use of the Ventcell transmission conditions. Another feature is the use of such algorithms in time-dependent problems in *space-time* domain decomposition that we have recently pioneered. This allows the use of different time steps in different parts of the computational domain and turns out to be particularly useful in porous media applications, where the amount of diffusion (permeability) varies abruptly, so that the evolution speed varies significantly from one part of the computational domain to another. Our new theme here are *Newton–multigrid solvers*, where the geometric multigrid solver is *tailored* to the specific problem under consideration and to the specific numerical method, with problem- and discretization-dependent restriction, prolongation, and smoothing. This in particular yields mass balance at each iteration step, a highly demanded feature in most of the target applications. The solver itself is then *adaptively steered* at each execution step by an a posteriori error estimate.

## 3.4. Reliability by a posteriori error control

The fourth part of our theoretical efforts goes towards guaranteeing the results obtained at the end of the numerical simulation. Here a key ingredient is the development of rigorous *a posteriori estimates* that make it possible to estimate in a fully computable way the error between the unknown exact solution and its numerical approximation. Our estimates also allow to distinguish the different *components* of the overall *error*, namely the errors coming from modeling, from the discretization scheme, from the nonlinear (Newton) solver, and from the linear algebraic (Krylov, domain decomposition, multigrid) solver. A new concept here is that of *local stopping criteria*, where all the error components are balanced locally within each computational mesh element. This naturally connects all parts of the numerical simulation process and gives rise to novel *fully adaptive algorithms*. We also theoretically address the question of convergence of the new fully adaptive algorithms. We identify theoretical conditions so that the error diminishes at each adaptive loop iteration by a contraction factor and we in particular derive a guaranteed error reduction factor in model cases. We shall also prove the numerical optimality of the derived algorithms in the sense that, up to a generic constant, the smallest possible computational effort to achieve the given accuracy is needed.

## 3.5. Safe and correct programming

Finally, we concentrate on the issue of computer implementation of scientific computing programs. Increasing complexity of algorithms for modern scientific computing makes it a major challenge to implement them in the traditional imperative languages popular in the community. As an alternative, the computer science community provides theoretically sound tools for *safe* and *correct programming*. We explore here the use of these tools to design generic solutions for the implementation of the class of scientific computing software that we deal with. Our focus ranges from high-level programming via *functional programming* with OCaml through safe and easy parallelism via *skeleton parallel programming* with SklML to proofs of correctness of numerical algorithms and programs via *mechanical proofs* with Coq.

# 4. Application Domains

## 4.1. Multiphase flows and transport of contaminants in the subsurface

- subsurface depollution after chemical leakage
- nuclear waste disposal in deep underground repositories
- flow in large scale discrete fracture networks
- production of oil and gas

## 4.2. Industrial risks in energy production

- Stokes and Navier–Stokes flows related to nuclear reactor operation
- reduced-order models for valves related to nuclear reactor operation
- plasticity and large deformations for mechanical components related to nuclear reactor operation
- seismic wave propagation for detection and protection
- electromagnetism for interfaces between dielectrics and negative metamaterials

## 4.3. Computational quantum chemistry

- guaranteed bounds for ground-state energy (eigenvalues) and ground-state density matrix (eigenvectors) in first-principle molecular simulation
- application to Laplace, Gross–Pitaevskii, Kohn–Sham, and Schrödinger models

# 5. Highlights of the Year

## 5.1. Highlights of the Year

Alexandre Ern co-edited with Daniele Di Pietro (Montpellier) and Luca Formaggia (Milano) a book on Numerical Methods for PDEs, SEMA SIMAI Springer Series, Vol. 15, Springer, 2018. ISBN 978-3-319-94675-7.

Many new results of the ERC GATIPOR project in the ERC GATIPOR Gallery.

# 6. New Software and Platforms

## 6.1. CELIA3D

KEYWORDS: Fluid mechanics - Multi-physics simulation

FUNCTIONAL DESCRIPTION: The CELIA3D code simulates the coupling between a compressible fluid flow and a deformable structure. The fluid is handled by a Finite Volume method on a structured Cartesian grid. The solid is handled by a Discrete Element method (Mka3d scheme). The solid overlaps the fluid grid and the coupling is carried out with immersed boundaries (cut cells) in a conservative way.

- Partners: Ecole des Ponts ParisTech - CEA
- Contact: Laurent Monasse
- URL: http://cermics.enpc.fr/~monassel/CELIA3D/

## 6.2. DiSk++

KEYWORDS: High order methods - Polyhedral meshes - C++

SCIENTIFIC DESCRIPTION: Discontinuous Skeletal methods approximate the solution of boundary-value problems by attaching discrete unknowns to mesh faces (hence the term skeletal) while allowing these discrete unknowns to be chosen independently on each mesh face (hence the term discontinuous). Cell-based unknowns, which can be eliminated locally by a Schur complement technique (also known as static condensation), are also used in the formulation. Salient examples of high-order Discontinuous Skeletal methods are Hybridizable Discontinuous Galerkin methods and the recently-devised Hybrid High-Order methods. Some major benefits of Discontinuous Skeletal methods are that their construction is dimension-independent and that they offer the possibility to use general meshes with polytopal cells and non-matching interfaces. The mathematical flexibility of Discontinuous Skeletal methods can be efficiently replicated in a numerical software: by using generic programming, the DiSk++ library offers an environment to allow a programmer to code mathematical problems in a way completely decoupled from the mesh dimension and the cell shape.

FUNCTIONAL DESCRIPTION: The software provides a numerical core to discretize partial differential equations arising from the engineering sciences (mechanical, thermal, diffusion). The discretization is based on the "Hybrid high-order" or "Discontinuous Skeletal" methods, which use as principal unknowns polynomials of arbitrary degree on each face of the mesh. An important feature of these methods is that they make it possible to treat general meshes composed of polyhedral cells. The DiSk ++ library, using generic programming techniques, makes it possible to write a code for a mathematical problem independently of the mesh. When a user writes the code for his problem using the basic operations offered by DiSk ++, that code can be executed without modifications on all types of mesh already supported by the library and those that will be added in the future.

- Author: Matteo Cicuttin
- Partner: CERMICS
- Contact: Matteo Cicuttin
- Publication: Implementation of Discontinuous Skeletal methods on arbitrary-dimensional, polytopal meshes using generic programming
- URL: https://github.com/wareHHOuse/diskpp

## 6.3. GENFIELD

KEYWORDS: Hydrogeology - Algorithm - Heterogeneity
SCIENTIFIC DESCRIPTION: GENFIELD implements a parallel version of the algorithm initially proposed by [E. Pardo-Iguzquiza and M. Chica-Olmo, Mathematical Geology, 25(2):177-217, 1993].
FUNCTIONAL DESCRIPTION: GENFIELD allows the generation of gaussian correlated fields. It is based on the circulant embedding method. Parallelism is implemented using MPI communications. GENFIELD is used in hydrogeology to model natural fields, like hydraulic conductivity or porosity fields.
NEWS OF THE YEAR: In 2018, we have performed scaling tests on ADA cluster. They have revealed that the symmetry of the phases required by this algorithm penalized a lot the parallel efficiency of GENFIELD (cf poster hal-01960444, version 1). We have decided to stop the development of GENFIELD. In 2018, we have developed a complete new software (cf BIL Inria ParaCIRCE) based on another algorithm initially proposed by [C. R. Dietrich and G. N. Newsam. A fast and exact method for multidimensional gaussian stochastic simulations. Water Resources Research, 29(8):2861-2869, 1993].

- Participants: Géraldine Pichot, Simon Legrand, Grégoire Lecourt, Jean-Raynald De Dreuzy and Jocelyne Erhel
- Contact: Géraldine Pichot
- Publications: GENFIELD: A parallel software for the generation of stationary Gaussian random fields - Algorithms for stationary Gaussian random field generation
- URL: https://gitlab.inria.fr/slegrand/Genfield_dev

## 6.4. Mka3d

KEYWORDS: Scientific computing - Elasticity - Elastodynamic equations
FUNCTIONAL DESCRIPTION: The Mka3d method simulates an elastic solid by discretizing the solid into rigid particles. An adequate choice of forces and torques between particles allows to recover the equations of elastodynamics.

- Partners: Ecole des Ponts ParisTech - CEA
- Contact: Laurent Monasse
- URL: http://cermics.enpc.fr/~monassel/Mka3D/

## 6.5. NEF-Draw

*Numerical Experiments involving Fractures -Visualisation*
KEYWORD: Fracture network
FUNCTIONAL DESCRIPTION: This software is a visualization tool of discrete fractured networks. It allows the visualization of the network geometry, the mesh of the network together with several quantities of interest (mesh quality, flow solution including wells) computed with the software NEF-Flow.
NEWS OF THE YEAR: This version includes Matlab vectorization of the operations which makes it possible to load flow solution on meshes with more than one million of fractures. It includes a text menu allowing the user to choose between different visualisation options (geometry, mesh together with the aspect ratio or together with the flow solution) A selective visualisation of fractures is also possible, loading only the fractures that carry most of the flow.

- Participant: Géraldine Pichot
- Contact: Géraldine Pichot
- URL: https://gitlab.inria.fr/gpichot/NEF

## 6.6. NEF-Flow

KEYWORDS: Hydrogeology - Numerical simulations - 3D

SCIENTIFIC DESCRIPTION: NEF-Flow is a Matlab software for the simulation of steady state single phase flow in Discrete Fracture Networks (DFNs) using the Mixed Hybrid Finite Element (MHFEM) method for conforming and non conforming discretizations.
FUNCTIONAL DESCRIPTION: The software NEF-Flow solves the problem of an incompressible fluid flowing through a network of fractures. The software is interfaced with different mesh generators, among which BLSURF from the GAMMA3 team. A mixed hybrid finite element method is implemented.
NEWS OF THE YEAR: The last version includes new feature: - wells, sink/source terms boundary conditions - Implementation of P1 non conforming finite elements - New data structures to save the information local to each fracture - New tests per fracture have been added to check the solution - Add wells and sink/sources boundary conditions in the function that check the solution.

- Participants: Géraldine Pichot, Jean-Raynald De Dreuzy and Jocelyne Erhel
- Contact: Géraldine Pichot
- Publication: A mixed hybrid Mortar method for solving flow in discrete fracture networks
- URL: https://gitlab.inria.fr/gpichot/NEF

## 6.7. ParaCirce

*Parallel Circulant Embedding*
KEYWORDS: 2D - 3D - Hydrogeology - Gaussian random fields - MPI
SCIENTIFIC DESCRIPTION: ParaCirce implements the algorithm proposed by [C. R. Dietrich and G. N. Newsam. A fast and exact method for multidimensional gaussian stochastic simulations. Water Resources Research, 29(8):2861-2869, 1993].
FUNCTIONAL DESCRIPTION: ParaCirce implements a parallel Circulant Embedding method for the generation in parallel of 2D or 3D Gaussian Random Fields (second order stationary).
NEWS OF THE YEAR: - MPI implementation - Dedicated C++ classes to allow a user-friendly and safe usage of the library - Efficient use of the external library RngStream (L'Ecuyer) for a guarantee of independent realizations and reproductibility. - Splitting of the domain along one direction. The repartition of the field is automatic or defined by the user. - Automatic computation of the padding

- Participants: Géraldine Pichot and Simon Legrand
- Contact: Géraldine Pichot
- URL: https://gitlab.inria.fr/slegrand/paracirce

## 6.8. PRune

*ParserRUNnEr*
KEYWORD: Test
FUNCTIONAL DESCRIPTION: Python tool to parse single or multi configurations parameters files and to automatically run a program and store the results in a predefined tree.

- Participants: Simon Legrand and Géraldine Pichot
- Contact: Simon Legrand

# 7. New Results

## 7.1. Unfitted hybrid-high-order methods

**Participants:** Alexandre Ern, Guillaume Delay.

Our team contributes actively to the development of hybrid high-order (HHO) methods. Such methods support polyhedral meshes with hanging nodes, but one requirement is that the mesh cells have planar faces. This is difficult when it comes to solving with high accuracy a problem posed on a domain with curved boundaries or a problem involving a curved interface separating two materials with different properties. One key idea to treat these problems is to use an unfitted mesh, so that the curved boundary or the curved interface freely cuts through the mesh cells. This greatly simplifies the meshing process, but at the same time poses the question on how the HHO method can address the approximation of functions that are not smooth within some mesh cells. The major idea in our approach, which is inspired from similar approaches developed in the context of the more classical finite element method, is to double the discrete unknowns attached to the cut mesh faces and to introduce a consistent Nitsche-type formulation to enforce either the boundary condition or the jump conditions across the interface in a weak manner. In this context, we started a collaboration with Erik Burman (University College London) and we elaborated in [20] the numerical analysis of HHO methods in an unfitted context; further analysis for Stokes and Helmholtz equations has started recently within the postdoc of Guillaume Delay and a collaboration on the subject with CEA is on the way.

## 7.2. An exponential time stepping scheme for the simulation of diffusion processes

**Participant:** Géraldine Pichot.

We present in [56] a new Monte Carlo algorithm to simulate diffusion processes in presence of discontinuous convective and diffusive terms. The algorithm is based on the knowledge of close form analytic expressions of the resolvents of the diffusion processes which are usually easier to obtain than close form analytic expressions of the density. In the particular case of diffusion processes with piecewise constant coefficients, known as Skew Diffusions, such close form expressions for the resolvent are available. Then we apply our algorithm to this particular case and we show that the approximate densities of the particles given by the algorithm replicate well the particularities of the true densities (discontinuities, bimodality, ...) Besides, numerical experiments show a quick convergence.

## 7.3. Localization of dual and distance norms

**Participants:** Martin Vohralík, Patrick Ciarlet Jr., Jan Blechta, Josef Málek.

Dual norms like the dual norm of the residual and the distance norm to the Sobolev space $H_0^1$ seem to be fundamentally global over the entire computational domain. In [23], together with P. Ciarlet, we prove, in extension of some older results, that they are both equivalent to the Hilbertian sums of their localizations over patches of elements. Together with J. Blechta and J. Málek, we extend in [45] this result from the space $H_0^1$ with Hilbertian structure to the Sobolev space $W_0^{1,p}$, with the exponent $p$ bigger than or equal to one, and to an arbitrary bounded linear functional on $W_0^{1,p}$.

## 7.4. Adaptivity with guaranteed error contraction

**Participants:** Martin Vohralík, Alexandre Ern, Patrik Daniel, Iain Smears.

In [26], we conceive novel adaptive refinement strategies which automatically decide between mesh refinement and polynomial degree increase. We numerically observe that the error decreases exponentially as a function of the number of degrees of freedom, for smooth as well as for singular numerical solutions. The salient feature of our approach is, however, that we ensure that the error on the next $hp$-refinement step will be reduced at least by a factor that is given. We then extend in [53] this result to the case where the underlying algebraic solver

is inexact. To the best of our knowledge, these results, obtained in the framework of the Ph.D. thesis of Patrik Daniel, is the first ever where such an error contraction bound is computable and guaranteed. Numerically, its precision turns out to be very high (overestimation by a factor very close to the optimal value of one). It immediately implies convergence of the adaptive method, and we would like to use it in the near future for optimality proofs.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

Three two-part contracts with EDF accompanying the PhD theses of Amina Benaceur, Nicolas Pignet, and Riccardo Milani.

Two two-part contract with CEA accompanying the PhD thesis of Frédéric Marazzato and the postdoc of Guillaume Delay.

Three-part contract Inria–EDF–Sciworks Technologies (from April 2017) on "Form-L for the formalization of constraints of complex systems". SERENA representants are Sébastien Furic and Pierre Weis.

AMIES NEF-PEPS1 (Dec. 2018–Feb. 2020) Collaboration with the joint laboratory LabCom fractory (ITASCA, Géosciences Rennes). SERENA representants are F. Clément, Sébastien Furic, Florent Hédin, M. Kern and G. Pichot (Coordinator).

Two-part contract with IFP Energies Nouvelles for co-supervision of the post-doc of G. Mallik.

# 9. Partnerships and Cooperations

## 9.1. Regional Initiatives

MILC (DMI RFSI, 2018–2019): "Mesure et Intégrale de Lebesgue en Coq", with LIPN (Université de Paris 13), and TOCCATA (Inria Saclay - Île-de-France). SERENA representants are François Clément and Vincent Martin (UTC).

GiS: scientific collaboration network between ten public institutions from the Paris (Ile-de-France) region, focused on natural resources and environment. The project-team SERENA is a member.

## 9.2. National Initiatives

### *9.2.1. ANR*

ANR HHOMM: "Hybrid high-order methods on polyhedral meshes", Theoretical foundations and applications (up to software development) for the recently-devised Hybrid high-order methods. Coordinated by D. Di Pietro, University of Montpellier. SERENA representant is A. Ern, period 2015–2019.

ANR DEDALES: "Algebraic and geometric domain decomposition for subsurface flow". The project aims at developing high performance software for the simulation of two phase flow in porous media. It specifically targets parallel computers where each node is itself composed of a large number of processing cores, such as are found in new generation many-core architectures.

The partners are HIEPACS, Laboratoire Analyse, Géométrie et Application, University Paris 13, Maison de la Simulation, and ANDRA. SERENA representants are M. Kern (grant leader) and M. Vohralík, period 2014–2018. The project ended in October 2018.

## 9.3. European Initiatives

### 9.3.1. FP7 & H2020 Projects

- EoCoE: “Energy Oriented Center of Excellence” This project is coordinated by Maison de la Simulation and gathers 23 partners from 13 countries to use the tremendous potential offered by the ever-growing computing infrastructure to foster and accelerate the European transition to a reliable low carbon energy supply using HPC (High Performance Computing). SERENA representant M. Kern, period 2015–2018.
- ERC GATIPOR: “Guaranteed fully adaptive algorithms with tailored inexact solvers for complex porous media flows”. The subject of this consolidator grant are new approaches to porous media multiphase flows: inexact Newton-multigrid solvers, local stopping criteria, adaptivity, and a posteriori error control. The goal is to guarantee the overall simulation error and to speed-up importantly the present-day simulations. SERENA representant is M. Vohralík (grant leader), period 2015–2020.
- PRACE: “Partnership for Advanced Computing in Europe” The mission of PRACE is to enable high-impact scientific discovery and engineering research and development across all disciplines to enhance European competitiveness for the benefit of society. PRACE has an extensive education and training effort for effective use of the Research Infrastructure. M. Kern is the French representative for training, and is in charge of the French node of the Prace training network, organizing 10-12 courses each year (period 2017-2019).

### 9.3.2. Collaborations in European Programs, Except FP7 & H2020

**OPENCPS**

Program: ITEA 3

Project acronym: OPENCPS

Project title: Open cyber-physical system model-driven certified development

Duration: Dec 2015–Dec 2018

Coordinator: Magnus Eek

Other partners: AB SKF, CEA, ELTE-Soft Kft., ESI Group, EDF, Wqua Simulation AB, Ericsson, IncQuery Labs Kft., KTH, Linköping University, RTE, SICS, SIREHNA, Saab AB, Sherpa Engineering, Siemens Industrial Torbumachinery AB, VTT Technical Research Center of Finland Ltd.

Abstract: Cyber-physical systems put increasing demands on reliability, usability, and flexibility while, at the same time, lead time and cost efficiency are essential for industry competitiveness. Tools and environments for model-based development of cyber-physical systems are becoming increasingly complex and critical for the industry: tool interoperability, vendor lock-ins, and tool life-cycle support are some of the challenges. The project focuses on interoperability between the standards Modelica/UML/FMI, improved execution speed of (co-)simulation, and certified code generation.

SERENA representants are Sébastien Furic and Pierre Weis.

## 9.4. International Initiatives

### 9.4.1. Inria International Partners

#### 9.4.1.1. Informal International Partners

Erik Burman, Professor at University College London, UK, unfitted methods.

Jean-Luc Guermond, Professor at Texas A&M University, USA, finite element methods.

Ulrich Rüde, Professor at Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany, multigrid methods.

Mary Wheeler, professor, University of Texas at Austin, USA, porous media applications.

Barbara Wohlmuth, Professor at Technical University of München, Germany, mixed finite element methods.

### *9.4.2. Participation in Other International Programs*

Alexandre Ern participated for two weeks in Jul 2018 as an invited scientist in the ESI Program on Numerical Analysis on Complex PDE in the Sciences, Vienna, Austria (https://www.esi.ac.at/activities/events/2018/numerical-analysis-of-complex-pde-models-in-the-sciences).

## 9.5. International Research Visitors

### *9.5.1. Visits of International Scientists*

Iain Smears, lecturer at University College London, March 26–30.

Thirupathi Gudi, Professor at Indian Institute of Science, Bangalore, India, January 15–February 28.

Jean-Luc Guermond, Professor at Texas A&M University, College Station, Texas, May 1–June 15.

Iain Smears, lecturer at University College London, June 18–27, and Christian Kreuzer, Professor at University Dortmund, June 18–29.

Carsten Carstensen, Professor at Humboldt University, Berlin, August 20–September 20.

Roland Becker, Professor at University of Pau, September 17–20.

Hend Ben Ameur, Professor at IPEST and member of ENIT-Lamsin, Tunis, Tunisia, November 19–30.

Théophile Chaumont-Frelet, junior researcher at Inria Sophia Antipolis, November 22–23.

*9.5.1.1. Internships*

Intissar Addali, 2nd year internship at ENSTA ParisTech, from May to Aug 2018, supervised by Karol Cascavita and Alexandre Ern.

### *9.5.2. Visits to International Teams*

*9.5.2.1. Research Stays Abroad*

Alexandre Ern visited the research group of Prof. Victor Calo, Curtin University, Perth, Australia, in November 2018.

Martin Vohralík was invited for two weeks stay to Charles University, Prague for collaboration with J. Málek, April 2018.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

Géraldine Pichot was the co-chair of the Computational Methods in Water Resources 2018 conference.

Ibtihel Ben Gharbia and Martin Vohralík have organized a 1-day workshop Journée contrat cadre IFP Energies Nouvelles/Inria.

*10.1.1.2. Member of the Organizing Committees*

Michel Kern was the member of the local organizing committee of the Computational Methods in Water Resources 2018 conference.

### 10.1.2. Scientific Events Selection

#### 10.1.2.1. Member of the Conference Program Committees

Alexandre Ern is a member of the Scientific Committee for the European Finite Element Fair.

Michel Kern was a member of the program committee for the JCAD 2018 (Journées Calcul et Données, Lyon October 2018).

Géraldine Pichot co-organized a mini-symposium, entitled "Numerical methods for processes in fractured media" at the InterPore 2018 - 10th Annual Meeting and Jubilee, New Orleans, United States, June 2018.

### 10.1.3. Journal

#### 10.1.3.1. Member of the Editorial Boards

François Clément was a member of the editorial board of Matapli until June 2018.

Alexandre Ern is a member of the editorial boards of SIAM Journal on Scientific Computing, ESAIM Mathematical Modelling and Numerical Analysis, IMA Journal of Numerical Analysis, and Computational Methods in Applied Mathematics.

Martin Vohralík is a member of the editorial boards of SIAM Journal on Numerical Analysis, Acta Polytechnica, and Applications of Mathematics.

#### 10.1.3.2. Reviewer - Reviewing Activities

Alexandre Ern served as reviewer for dozens of papers in different journals.

Michel Kern was a reviewer for OGST, BIT, Mathematics and Computers in Simulation, Computers and Geosciences.

Martin Vohralík served as reviewer for dozens of papers in different journals.

### 10.1.4. Invited Talks

Alexandre Ern gave an invited lecture at the BAIL 2018 conference in Glasgow (*International conference on Boundary and Interior Layers*).

Géraldine Pichot was an invited speaker at the InterPore 2018 - 10th Annual Meeting and Jubilee, New Orleans, United States, June 2018 and at the Workshop Reactive Flows in Deformable, Complex Media, Oberwolfach, Germany, August 2018.

Martin Vohralík was an invited speaker at the workshop FEEC and High Order Methods, Oslo, Norway, the Seventh Conference on Finite Difference Methods: Theory and Applications, Lozenetz, Bulgaria, and the Workshop Reactive Flows in Deformable, Complex Media, Oberwolfach, Germany, August 2018.

### 10.1.5. Leadership within the Scientific Community

Alexandre Ern is the leader of the Master Mathématiques et applications, Ecole nationale des ponts et chaussées.

M. Kern is a member of the Scientific Committee of Orap (ORganisation Associative du Parallélisme), of the steering committee of Géosciences franciliennes of the Scientific Board of GDR Calcul, and of the jury and executive board of Label C3I.

M. Vohralík is a member of the steering committees of Géosciences franciliennes and Summer schools CEA–EDF–Inria.

M. Vohralík is in charge of the topic "Numerical schemes, mesh generation algorithms, and error control" in the ANDRA, BRGM, CEA, EDF, IFP Energies Nouvelles, and Total working group on *High-Performance Numerical Simulation in the Geosciences* (identification of common challenges and collaboration opportunities).

### 10.1.6. Scientific Expertise

M. Kern is a reviewer for the German Supercomputing Center JARA program.

### *10.1.7. Research Administration*

François Clément is a member of the *Comité local d'hygiène, de sécurité et des conditions de travail* of the Inria Research Center of Paris.

François Clément was the *AMIES facilitator* of the Inria Research Center of Paris until June 2018.

M. Kern is Deputy Director of Maison de la Simulation, a joint project between CEA, CNRS, Inria, Université de Paris 11, and Université de Versailles, focused on applications of high end computing.

M. Kern is a member of the *Comité de site* of the Inria center of Paris.

Géraldine Pichot was a member of the *Comité local d'hygiène, de sécurité et des conditions de travail* of the Inria Research Center of Paris until Sep 2018.

Géraldine Pichot was a member of the *Commission de développement technologique* of the Inria Research Center of Paris until March 2018.

Martin Vohralík is a member of the Inria Paris *Committee on scientific positions* (evaluation of applications for Ph.D. theses (CORDI-S), post-docs, and "délégations").

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

Licence : Alexandre Ern, Optimal Control, 20h, L3, Ecole Polytechnique, France.

Licence : Alexandre Ern, Partial differential equations, 10h, L3, Ecole nationale des ponts et chaussées, France.

Master : Alexandre Ern, Discontinuous Galerkin methods, 20h, M2, Sorbonne University, France.

Master: Michel Kern, Inverse Problems, 26h, M1, Mines-ParisTech, France

Master: Michel Kern, Advanced Numerical Analysis, 30h, M2, Institut Galilée, Université Paris 13, France

Master: Michel Kern, Subsurface flows, 30h (with E. Mouche), M2, Université Paris Saclay, France

Master: Martin Vohralík, A posteriori error estimates for efficiency and error control in numerical simulations, 36h, M2, Charles University, Prague, Czech Republic.

### *10.2.2. Supervision*

PhD: Amina Benaceur, Model reduction for nonlinear thermics and mechanics, 21 Dec 2018, Alexandre Ern.

PhD: Karol Cascavita, Hybrid discretization methods for Signorini contact and Bingham flow problems, 18 Dec 2018, Alexandre Ern and Xavier Chateau.

PhD in progress: Jad Dabaghi, A posteriori error estimates and adaptive stopping criteria for formulations with complementarity constraints, 01 November 2015, Martin Vohralík and Vincent Martin.

PhD in progress: Patrik Daniel, Adaptive $hp$-finite elements with guaranteed error contraction and inexact multilevel solvers, 01 October 2015, Martin Vohralík and Alexandre Ern.

PhD in progress: Frédéric Marazzato, Discrete element methods for fracture and fragmentation, 01 October 2016, Alexandre Ern.

PhD in progress: Riccardo Milani, Compatible Discrete Operator schemes for Navier–Stokes equations, 01 October 2017, Alexandre Ern.

PhD in progress: Ani Miraci, Robust a posteriori error control and adaptivity with inexact solvers, 01 October 2017, Martin Vohralík and Alexandre Ern.

PhD in progress: Nicolas Pignet, Hybrid High-Order methods for nonlinear mechanics, 01 November 2016, Alexandre Ern.

### *10.2.3. Juries*

Alexandre Ern, Referee, PhD A. Bensalah, ENSTA ParisTech, Jul 2018.

Alexandre Ern, Examiner, PhD G. Morel, Sorbonne University, Sep 2018.

Alexandre Ern, Referee, PhD P. Vega, University of Concepcion, Chile, Nov 2018.

Alexandre Ern, Referee, PhD G. Pennesi, Politecnico Milano, Italy, Dec 2018.

Michel Kern, Examiner, HDR J. Carrayrou, University of Strasbourg, 21 March 2018.

Martin Vohralík, Referee, PhD O. Gorynina, Université de Bourgogne Franche-Comté, Besançon, Feb 2018.

Martin Vohralík, Examiner, PhD M. Botti, Université de Montpellier, Nov 2018.

## 10.3. Popularization

### *10.3.1. Interventions*

- M. Kern: présentation “Des mathématiques pour modéliser et simuler le monde”, Lycée des Pierres–Vives (Terminale S), Carrières–sur–Seine, May 2018.
- M. Vohralík: “Advancing scientific knowledge together to support innovation”, dissemination video, IFP Energies Nouvelles. Available here. February 2018.

Major publications by the team in recent years:

# 11. Bibliography

## Major publications by the team in recent years

[1] S. BOLDO, F. CLÉMENT, J.-C. FILLIÂTRE, M. MAYERO, G. MELQUIOND, P. WEIS.*Wave equation numerical resolution: a comprehensive mechanized proof of a C program*, in "Journal of Automated Reasoning", April 2013, vol. 50, n$^{o}$ 4, p. 423–456, http://dx.doi.org/10.1007/s10817-012-9255-4

[2] S. BOLDO, F. CLÉMENT, J.-C. FILLIÂTRE, M. MAYERO, G. MELQUIOND, P. WEIS.*Trusting computations: A mechanized proof from partial differential equations to actual program*, in "Computers and Mathematics with Applications", August 2014, vol. 68, n$^{o}$ 3, p. 325–352, http://dx.doi.org/10.1016/j.camwa.2014.06.004

[3] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Guaranteed and robust a posteriori bounds for Laplace eigenvalues and eigenvectors: conforming approximations*, in "SIAM J. Numer. Anal.", 2017, vol. 55, n$^{o}$ 5, p. 2228–2254, http://dx.doi.org/10.1137/15M1038633

[4] D. A. DI PIETRO, A. ERN.*A hybrid high-order locking-free method for linear elasticity on general meshes*, in "Comput. Methods Appl. Mech. Engrg.", 2015, vol. 283, p. 1–21, http://dx.doi.org/10.1016/j.cma.2014.09.009

[5] A. ERN, J.-L. GUERMOND.*Finite element quasi-interpolation and best approximation*, in "ESAIM Math. Model. Numer. Anal.", 2017, vol. 51, n$^{o}$ 4, p. 1367–1385, https://doi.org/10.1051/m2an/2016066

[6] A. ERN, M. VOHRALÍK.*Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations*, in "SIAM J. Numer. Anal.", 2015, vol. 53, n$^{o}$ 2, p. 1058–1081, http://dx.doi.org/10.1137/130950100

[7] T.-T.-P. HOANG, J. JAFFRÉ, C. JAPHET, M. KERN, J. E. ROBERTS.*Space-time domain decomposition methods for diffusion problems in mixed formulations*, in "SIAM J. Numer. Anal.", 2013, vol. 51, n$^{o}$ 6, p. 3532–3559, http://dx.doi.org/10.1137/130914401

[8] T.-T.-P. HOANG, C. JAPHET, M. KERN, J. E. ROBERTS.*Space-time domain decomposition for reduced fracture models in mixed formulation*, in "SIAM J. Numer. Anal.", 2016, vol. 54, n$^{o}$ 1, p. 288–316, http://dx.doi.org/10.1137/15M1009651

[9] A. LEJAY, G. PICHOT.*Simulating diffusion processes in discontinuous media: a numerical scheme with constant time steps*, in "J. Comput. Phys.", 2012, vol. 231, n$^{o}$ 21, p. 7299–7314, http://dx.doi.org/10.1016/j.jcp.2012.07.011

[10] G. PICHOT, J. ERHEL, J.-R. DE DREUZY.*A generalized mixed hybrid mortar method for solving flow in stochastic discrete fracture networks*, in "SIAM J. Sci. Comput.", 2012, vol. 34, n$^{o}$ 1, p. B86–B105, http://dx.doi.org/10.1137/100804383

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] A. BENACEUR.*Réduction de modèles en thermique et mécanique non-linéaires*, Université Paris-Est Marne la Vallée, December 2018, https://hal.archives-ouvertes.fr/tel-01958278

### Articles in International Peer-Reviewed Journal

[12] M. ABBAS, A. ERN, N. PIGNET.*Hybrid High-Order methods for finite deformations of hyperelastic materials*, in "Computational Mechanics", January 2018, vol. 62, n$^{o}$ 4, p. 909-928 [*DOI :* 10.1007/s00466-018-1538-0], https://hal.archives-ouvertes.fr/hal-01575370

[13] M. ABBAS, A. ERN, N. PIGNET.*A Hybrid High-Order method for incremental associative plasticity with small deformations*, in "Computer Methods in Applied Mechanics and Engineering", April 2019, vol. 346, p. 891-912, https://arxiv.org/abs/1804.06129 [*DOI :* 10.1016/j.cma.2018.08.037], https://hal.archives-ouvertes.fr/hal-01768411

[14] E. AHMED, A. BEN ABDA.*The sub-Cauchy Stokes Problem: Solvability Issues and Lagrange Multiplier Methods with Artificial Boundary Conditions*, in "Journal of Computational and Applied Mathematics", January 2018, https://hal.archives-ouvertes.fr/hal-01467425

[15] S. ALI HASSAN, C. JAPHET, M. KERN, M. VOHRALÍK.*A posteriori stopping criteria for optimized Schwarz domain decomposition algorithms in mixed formulations*, in "Computational Methods in Applied Mathematics", June 2018, vol. 18, n$^{o}$ 3, p. 495-519 [*DOI :* 10.1515/cmam-2018-0010], https://hal.inria.fr/hal-01529532

[16] S. ALI HASSAN, C. JAPHET, M. VOHRALÍK.*A posteriori stopping criteria for space-time domain decomposition for the heat equation in mixed formulations* , in "Electronic Transactions on Numerical Analysis", 2018, vol. 49, p. 151–181, [*DOI :* 10.1553/etna_vol49s151], https://hal.inria.fr/hal-01586862

[17] L. AMIR, M. KERN.*Preconditioning a coupled model for reactive transport in porous media*, in "International Journal of Numerical Analysis and Modeling", 2018, vol. 16, n$^{o}$ 1, p. 1-30, https://arxiv.org/abs/1710.01483 , https://hal.inria.fr/hal-01327307

[18] H. BEN AMEUR, G. CHAVENT, F. CHEIKH, F. CLÉMENT, V. MARTIN, J. E. ROBERTS.*First-order indicators for the estimation of discrete fractures in porous media*, in "Inverse Problems in Science and Engineering", 2018, vol. 26, n$^{o}$ 1, p. 1–32, https://arxiv.org/abs/1602.08304 [*DOI :* 10.1080/17415977.2017.1290087], https://hal.inria.fr/hal-01279503

[19] A. BENACEUR, V. EHRLACHER, A. ERN, S. MEUNIER.*A progressive reduced basis/empirical interpolation method for nonlinear parabolic problems*, in "SIAM Journal on Scientific Computing", 2018, vol. 40, n$^{o}$ 5, p. A2930-A2955, https://arxiv.org/abs/1710.00511 , https://hal.archives-ouvertes.fr/hal-01599304

[20] E. BURMAN, A. ERN.*An unfitted Hybrid High-Order method for elliptic interface problems*, in "SIAM Journal on Numerical Analysis", 2018, vol. 56, n$^{o}$ 3, p. 1525-1546, https://hal.archives-ouvertes.fr/hal-01625421

[21] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Guaranteed and robust a posteriori bounds for Laplace eigenvalues and eigenvectors: a unified framework*, in "Numerische Mathematik", July 2018, vol. 140, n$^{o}$ 4, p. 1033-1079 [*DOI :* 10.1007/s00211-018-0984-0], https://hal.inria.fr/hal-01483461

[22] K. L. CASCAVITA, J. BLEYER, X. CHATEAU, A. ERN.*Hybrid discretization methods with adaptive yield surface detection for Bingham pipe flows*, in "Journal of Scientific Computing", June 2018, vol. 77, n$^{o}$ 3, p. 1424-1443 [*DOI :* 10.1007/s10915-018-0745-3], https://hal.archives-ouvertes.fr/hal-01698983

[23] P. CIARLET, M. VOHRALÍK.*Localization of global norms and robust a posteriori error control for transmission problems with sign-changing coefficients*, in "Modelisation Mathématique et Analyse Numérique", December 2018, vol. 52, n$^{o}$ 5, p. 2037-2064 [*DOI :* 10.1051/m2an/2018034], https://hal.inria.fr/hal-01148476

[24] M. CICUTTIN, D. A. DI PIETRO, A. ERN.*Implementation of Discontinuous Skeletal methods on arbitrary-dimensional, polytopal meshes using generic programming*, in "Journal of Computational and Applied Mathematics", 2018, vol. 344, p. 852–874 [*DOI :* 10.1016/j.cam.2017.09.017], https://hal.archives-ouvertes.fr/hal-01429292

[25] 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

[26] P. DANIEL, A. ERN, I. SMEARS, M. VOHRALÍK.*An adaptive hp-refinement strategy with computable guaranteed bound on the error reduction factor*, in "Computers and Mathematics with Applications", September 2018, vol. 76, n$^{o}$ 5, p. 967-983, https://arxiv.org/abs/1712.09821 [*DOI :* 10.1016/j.camwa.2018.05.034], https://hal.inria.fr/hal-01666763

[27] A. ERN, J.-L. GUERMOND.*Abstract nonconforming error estimates and application to boundary penalty methods for diffusion equations and time-harmonic Maxwell's equations*, in "Computational Methods in Applied Mathematics", 2018, vol. 18, n$^{o}$ 3, p. 451-475 [*DOI :* 10.1515/cmam-2017-0058], https://hal.archives-ouvertes.fr/hal-01563594

[28] A. ERN, I. SMEARS, M. VOHRALÍK.*Equilibrated flux a posteriori error estimates in $L^2(H^1)$-norms for high-order discretizations of parabolic problems*, in "IMA Journal of Numerical Analysis", June 2018 [*DOI :* 10.1093/imanum/dry035], https://hal.inria.fr/hal-01489721

[29] M. KÖPPEL, V. MARTIN, J. JAFFRÉ, J. E. ROBERTS.*A Lagrange multiplier method for a discrete fracture model for flow in porous media*, in "Computational Geosciences", September 2018, https://hal.archives-ouvertes.fr/hal-01700663

[30] F. MARAZZATO, A. ERN, C. MARIOTTI, L. MONASSE.*An explicit pseudo-energy conserving time-integration scheme for Hamiltonian dynamics*, in "Computer Methods in Applied Mechanics and Engineering", 2019 [*DOI :* 10.1016/J.CMA.2019.01.013], https://hal-enpc.archives-ouvertes.fr/hal-01661608

[31] J. PAPEŽ, Z. STRAKOŠ, M. VOHRALÍK.*Estimating and localizing the algebraic and total numerical errors using flux reconstructions*, in "Numerische Mathematik", February 2018, vol. 138, n$^{o}$ 3, p. 681-721 [*DOI :* 10.1007/S00211-017-0915-5], https://hal.inria.fr/hal-01312430

[32] M. VOHRALÍK, S. YOUSEF.*A simple a posteriori estimate on general polytopal meshes with applications to complex porous media flows*, in "Computer Methods in Applied Mechanics and Engineering", April 2018, vol. 331, p. 728-760 [*DOI :* 10.1016/J.CMA.2017.11.027], https://hal.archives-ouvertes.fr/hal-01532195

[33] M. ČERMÁK, F. HECHT, Z. TANG, M. VOHRALÍK.*Adaptive inexact iterative algorithms based on polynomial-degree-robust a posteriori estimates for the Stokes problem*, in "Numerische Mathematik", February 2018, vol. 138, n$^{o}$ 4, p. 1027-1065 [*DOI :* 10.1007/S00211-017-0925-3], https://hal.inria.fr/hal-01097662

### Invited Conferences

[34] H. BARUCQ, H. CALANDRA, G. CHAVENT, F. FAUCHER.*Stability and convergence analysis for seismic depth imaging using FWI*, in "Reconstruction Methods for Inverse Problems", Rome, Italy, Reconstruction Methods for Inverse Problems, May 2018, https://hal.archives-ouvertes.fr/hal-01807980

[35] G. PICHOT, P. LAUG, J. ERHEL, R. LE GOC, C. DARCEL, P. DAVY, J.-R. DE DREUZY.*Flow simulations in geology-based Discrete Fracture Networks*, in "2018 - Reactive Flows in Deformable, Complex Media", Oberwolfach, Germany, August 2018, p. 1-3, https://hal.inria.fr/hal-01900605

[36] G. PICHOT, P. LAUG, R. LE GOC, C. DARCEL, P. DAVY, J.-R. DE DREUZY.*Computation of flow properties of large scale fractured media*, in "InterPore 2018 - 10th Annual Meeting and Jubilee", New Orleans, United States, June 2018, https://hal.inria.fr/hal-01900599

### International Conferences with Proceedings

[37] S. LOPEZ, R. MASSON, L. BEAUDE, N. BIRGLE, K. BRENNER, M. KERN, F. SMAÏ, F. XING.*Geothermal Modeling in Complex Geological Systems with the ComPASS Code*, in "Stanford Geothermal Workshop 2018 - 43rd Workshop on Geothermal Reservoir Engineering", Stanford, United States, Stanford University, February 2018, https://hal-brgm.archives-ouvertes.fr/hal-01667379

### Conferences without Proceedings

[38] P. LAUG, G. PICHOT, R. LE GOC, C. DARCEL, P. DAVY.*Automatic meshing of Discrete Fracture Networks*, in "Computational Methods in Water Resources XXII (CMWR 2018)", Saint-Malo, France, June 2018, https://hal.inria.fr/hal-01896927

[39] P. LAUG, G. PICHOT.*Simulations in large tridimensional Discrete Fracture Networks (DFN): I. Geometric modeling and mesh generation*, in "MASCOT 2018 - 15th IMACS/ISGG meeting on applied scientific computing and tools", Rome, Italy, October 2018, p. 1-2, https://hal.inria.fr/hal-01896881

[40] A. LEJAY, G. PICHOT, L. LENÔTRE.*Diffusion processes in discontinuous media: numerical algorithms and benchmark tests*, in "Workshop Validation approaches for multiscale porous media models", Nottingham, United Kingdom, July 2018, https://hal.inria.fr/hal-01900609

[41] S. LOPEZ, R. MASSON, F. XING, L. BEAUDE, F. SMAÏ, M. KERN, A. ARMANDINE LES LANDES, G. AMIEZI, K. BRENNER, G. COURRIOUX, S. CARITG-MONNOT.*Modélisation hydrothermale des systèmes géothermiques profonds fracturés avec le code ComPASS*, in "26ème Réunion des Sciences de la Terre - RST", Lille, France, October 2018, https://hal-brgm.archives-ouvertes.fr/hal-01890182

[42] G. PICHOT, P. LAUG, J. ERHEL, R. LE GOC, C. DARCEL, P. DAVY, J.-R. DE DREUZY.*Simulations in large tridimensional Discrete Fracture Networks (DFN): II. Flow simulations*, in "MASCOT 2018 -15th IMACS/ISGG meeting on applied scientific computing and tools", Rome, Italy, October 2018, https://hal.inria.fr/hal-01896900

## Other Publications

[43] M. ABBAS, A. ERN, N. PIGNET.*A Hybrid High-Order method for finite elastoplastic deformations within a logarithmic strain framework*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01978385

[44] I. BEN GHARBIA, J. DABAGHI, V. MARTIN, M. VOHRALÍK.*A posteriori error estimates and adaptive stopping criteria for a compositional two-phase flow with nonlinear complementarity constraints*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01919067

[45] J. BLECHTA, J. MÁLEK, M. VOHRALÍK.*Localization of the $W^{-1,q}$ norm for local a posteriori efficiency*, July 2018, working paper or preprint, https://hal.inria.fr/hal-01332481

[46] T. BOIVEAU, V. EHRLACHER, A. ERN, A. NOUY.*Low-rank approximation of linear parabolic equations by space-time tensor Galerkin methods*, October 2018, https://arxiv.org/abs/1712.07256 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01668316

[47] V. M. CALO, M. CICUTTIN, Q. DENG, A. ERN.*Spectral approximation of elliptic operators by the Hybrid High-Order method*, July 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01628698

[48] E. CANCÈS, G. DUSSON, Y. MADAY, B. STAMM, M. VOHRALÍK.*Post-processing of the planewave approximation of Schrödinger equations. Part I: linear operators*, November 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01908039

[49] 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

[50] J.-P. CHANCELIER, S. FURIC, P. WEIS.*Translating Simulink Models to Modelica using the Nsp Platform*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01948681

[51] M. CICUTTIN, A. ERN, T. GUDI.*Discontinuous-Skeletal methods with linear and quadratic reconstructions for the elliptic obstacle problem*, February 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01718883

[52] J. DABAGHI, V. MARTIN, M. VOHRALÍK.*Adaptive inexact semismooth Newton methods for the contact problem between two membranes*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01666845

[53] P. DANIEL, A. ERN, M. VOHRALÍK.*An adaptive hp-refinement strategy with inexact solvers and computable guaranteed bound on the error reduction factor*, November 2018, working paper or preprint, https://hal.inria.fr/hal-01931448

[54] A. ERN, J.-L. GUERMOND.*Quasi-optimal nonconforming approximation of elliptic PDES with contrasted coefficients and minimal regularity*, December 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01964299

[55] A. ERN, M. VOHRALÍK.*Stable broken H1 and H(div) polynomial extensions for polynomial-degree-robust potential and flux reconstruction in three space dimensions*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01422204

[56] A. LEJAY, L. LENÔTRE, G. PICHOT.*An exponential timestepping algorithm for diffusion with discontinuous coefficients*, June 2018, working paper or preprint, https://hal.inria.fr/hal-01806465

[57] G. MALLIK, M. VOHRALÍK, S. YOUSEF.*Goal-oriented a posteriori error estimation for conforming and nonconforming approximations with inexact solvers*, December 2018, working paper or preprint, https://hal.inria.fr/hal-01964733

[58] G. PICHOT, S. LEGRAND, J. ERHEL, M. OUMOUNI.*GENFIELD: A parallel software for the generation of stationary Gaussian random fields*, May 2018, 1, InterPore 2018 - 10th Annual Meeting and Jubilee, Poster, https://hal.inria.fr/hal-01960444

[59] M. RIAHI, H. BEN AMEUR, J. JAFFRÉ, R. BOUHLILA.*Refinement indicators for estimating hydrogeologic parameters*, January 2018, working paper or preprint, https://hal.inria.fr/hal-01674486

[60] I. SMEARS, M. VOHRALÍK.*Simple and robust equilibrated flux a posteriori estimates for singularly perturbed reaction-diffusion problems*, 2018, working paper or preprint, https://hal.inria.fr/hal-01956180

# Project-Team SIERRA

## Statistical Machine Learning and Parsimony

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Optimization, machine learning and statistical methods**

# Table of contents

## Project-Team SIERRA

*Creation of the Team: 2011 January 01, updated into Project-Team: 2012 January 01*

**Keywords:**

**Computer Science and Digital Science:**

A1.2.8. - Network security
A3.4. - Machine learning and statistics
A5.4. - Computer vision
A6.2. - Scientific computing, Numerical Analysis & Optimization
A7.1. - Algorithms
A8.2. - Optimization
A9.2. - Machine learning

**Other Research Topics and Application Domains:**

B9.5.6. - Data science

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Francis Bach [Team leader, Inria, Senior Researcher, HDR]
Alexandre d'Aspremont [CNRS, Senior Researcher, HDR]
Pierre Gaillard [Inria, Researcher]
Alessandro Rudi [Inria, Starting Research Position]

**Technical Staff**

Loïc Estève [Inria, from Apr 2018]
Hadrien Hendrikx [Inria, from Apr 2018 until Sep 2018]

**PhD Students**

Remi Leblond [Inria, Researcher, until Aug 2018]
Dmitry Babichev [Inria]
Mathieu Barré [Ecole Normale Supérieure Paris, from Sep 2018]
Raphaël Berthier [Inria, from Oct 2018]
Anaël Bonneton [Ecole Normale Supérieure Paris]
Margaux Brégère [EDF]
Alexandre Défossez [Facebook]
Radu Alexandru Dragomir [Ecole polytechnique, from Sep 2018]
Thomas Kerdreux [Ecole polytechnique]
Gregoire Mialon [Inria, from Oct 2018]
Loucas Pillaud Vivien [Ministère de l'Ecologie, de l'Energie, du Développement durable et de la Mer]
Antoine Recanati [CNRS, until Sep 2018]
Damien Scieur [Inria, until Aug 2018]
Tatiana Shpakova [Inria]
Alex Nowak Vila [Inria, from Oct 2018]

**Post-Doctoral Fellows**

Lenaic Chizat [Inria, until Nov 2018]
Pierre Yves Massé [Université Technique de Prague, from Apr 2018]
Dmitrii Ostrovskii [Inria, from Feb 2018]
Adrien Taylor [Inria]

**Visiting Scientists**

Vijaya Bollapragada [Northwestern University, from Apr 2018 until Jul 2018]
Aaron Defazio [Facebook Research, until Feb 2018]
Gauthier Gidel [University of Montreal, Jan 2018]
Achintya Kundu [Ecole d'ingénieurs, from Jun 2018 until Aug 2018]
Gregoire Mialon [Inria, Sep 2018]
Sharan Vaswani [University of British Columbia, from Apr 2018 until Jul 2018]
Simon Lacoste-Julien [University of Montreal, Aug 2018]

**Administrative Assistants**

Helene Bessin Rousseau [Inria, from Mar 2018]
Sabrine Boumizy [Inria, until Feb 2018]
Sandrine Verges [Inria, until Jan 2018]

# 2. Overall Objectives

## 2.1. Statement

Machine learning is a recent scientific domain, positioned between applied mathematics, statistics and computer science. Its goals are the optimization, control, and modelisation of complex systems from examples. It applies to data from numerous engineering and scientific fields (e.g., vision, bioinformatics, neuroscience, audio processing, text processing, economy, finance, etc.), the ultimate goal being to derive general theories and algorithms allowing advances in each of these domains. Machine learning is characterized by the high quality and quantity of the exchanges between theory, algorithms and applications: interesting theoretical problems almost always emerge from applications, while theoretical analysis allows the understanding of why and when popular or successful algorithms do or do not work, and leads to proposing significant improvements.

Our academic positioning is exactly at the intersection between these three aspects—algorithms, theory and applications—and our main research goal is to make the link between theory and algorithms, and between algorithms and high-impact applications in various engineering and scientific fields, in particular computer vision, bioinformatics, audio processing, text processing and neuro-imaging.

Machine learning is now a vast field of research and the team focuses on the following aspects: supervised learning (kernel methods, calibration), unsupervised learning (matrix factorization, statistical tests), parsimony (structured sparsity, theory and algorithms), and optimization (convex optimization, bandit learning). These four research axes are strongly interdependent, and the interplay between them is key to successful practical applications.

# 3. Research Program

## 3.1. Supervised Learning

This part of our research focuses on methods where, given a set of examples of input/output pairs, the goal is to predict the output for a new input, with research on kernel methods, calibration methods, and multi-task learning.

## 3.2. Unsupervised Learning

We focus here on methods where no output is given and the goal is to find structure of certain known types (e.g., discrete or low-dimensional) in the data, with a focus on matrix factorization, statistical tests, dimension reduction, and semi-supervised learning.

## 3.3. Parsimony

The concept of parsimony is central to many areas of science. In the context of statistical machine learning, this takes the form of variable or feature selection. The team focuses primarily on structured sparsity, with theoretical and algorithmic contributions.

## 3.4. Optimization

Optimization in all its forms is central to machine learning, as many of its theoretical frameworks are based at least in part on empirical risk minimization. The team focuses primarily on convex and bandit optimization, with a particular focus on large-scale optimization.

# 4. Application Domains

## 4.1. Applications for Machine Learning

Machine learning research can be conducted from two main perspectives: the first one, which has been dominant in the last 30 years, is to design learning algorithms and theories which are as generic as possible, the goal being to make as few assumptions as possible regarding the problems to be solved and to let data speak for themselves. This has led to many interesting methodological developments and successful applications. However, we believe that this strategy has reached its limit for many application domains, such as computer vision, bioinformatics, neuro-imaging, text and audio processing, which leads to the second perspective our team is built on: Research in machine learning theory and algorithms should be driven by interdisciplinary collaborations, so that specific prior knowledge may be properly introduced into the learning process, in particular with the following fields:

- Computer vision: object recognition, object detection, image segmentation, image/video processing, computational photography. In collaboration with the Willow project-team.
- Bioinformatics: cancer diagnosis, protein function prediction, virtual screening. In collaboration with Institut Curie.
- Text processing: document collection modeling, language models.
- Audio processing: source separation, speech/music processing.
- Neuro-imaging: brain-computer interface (fMRI, EEG, MEG).

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Awards*

Francis Bach, Lagrange Prize in Continuous Optimization, Society for Industrial and Applied Mathematics 2018

Francis Bach, Best Paper Award, NeurIPS 2018.

Francis Bach included in the report *Highly cited researchers, year 2018*, Clarivate Analytics, 2018

Nicolas Flammarion, PhD thesis award in the *Programme Gaspard Monge*, Fondation Mathématique Jacques Hadamard, 2018.

Adrien Taylor, Tucker Prize (finalist) 2018 (dissertation prize by the Math- ematical Optimization Society for 2015-2017).

Adrien Taylor, IBM/FNRS innovation award 2018 (dissertation prize for original contributions to informatics).

Adrien Taylor, Icteam thesis award 2018 (dissertation award by the icteam institute of UCLouvain, Belgium).

Adrien Taylor, Best paper award 2018 from the journal Optimization Letters for the paper *On the worst-case complexity of the gradient method with exact line search for smooth strongly convex functions*, Etienne De Klerk, François Glineur, Adrien Taylor. journal=.

# 6. New Software and Platforms

## 6.1. ProxASAGA

KEYWORD: Optimization
FUNCTIONAL DESCRIPTION: A C++/Python code implementing the methods in the paper "Breaking the Nonsmooth Barrier: A Scalable Parallel Method for Composite Optimization", F. Pedregosa, R. Leblond and S. Lacoste-Julien, Advances in Neural Information Processing Systems (NIPS) 2017. Due to their simplicity and excellent performance, parallel asynchronous variants of stochastic gradient descent have become popular methods to solve a wide range of large-scale optimization problems on multi-core architectures. Yet, despite their practical success, support for nonsmooth objectives is still lacking, making them unsuitable for many problems of interest in machine learning, such as the Lasso, group Lasso or empirical risk minimization with convex constraints. In this work, we propose and analyze ProxASAGA, a fully asynchronous sparse method inspired by SAGA, a variance reduced incremental gradient algorithm. The proposed method is easy to implement and significantly outperforms the state of the art on several nonsmooth, large-scale problems. We prove that our method achieves a theoretical linear speedup with respect to the sequential version under assumptions on the sparsity of gradients and block-separability of the proximal term. Empirical benchmarks on a multi-core architecture illustrate practical speedups of up to 12x on a 20-core machine.

- Contact: Fabian Pedregosa
- URL: https://github.com/fabianp/ProxASAGA

## 6.2. object-states-action

KEYWORD: Computer vision
FUNCTIONAL DESCRIPTION: Code for the paper Joint Discovery of Object States and Manipulation Actions, ICCV 2017: Many human activities involve object manipulations aiming to modify the object state. Examples of common state changes include full/empty bottle, open/closed door, and attached/detached car wheel. In this work, we seek to automatically discover the states of objects and the associated manipulation actions. Given a set of videos for a particular task, we propose a joint model that learns to identify object states and to localize state-modifying actions. Our model is formulated as a discriminative clustering cost with constraints. We assume a consistent temporal order for the changes in object states and manipulation actions, and introduce new optimization techniques to learn model parameters without additional supervision. We demonstrate successful discovery of seven manipulation actions and corresponding object states on a new dataset of videos depicting real-life object manipulations. We show that our joint formulation results in an improvement of object state discovery by action recognition and vice versa.

- Participants: Jean-Baptiste Alayrac, Josef Sivic, Ivan Laptev and Simon Lacoste-Julien
- Contact: Jean-Baptiste Alayrac
- Publication: Joint Discovery of Object States and Manipulation Actions
- URL: https://github.com/jalayrac/object-states-action

# 7. New Results

## 7.1. On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport

Many tasks in machine learning and signal processing can be solved by minimizing a convex function of a measure. This includes sparse spikes deconvolution or training a neural network with a single hidden layer. For these problems, in [25] we study a simple minimization method: the unknown measure is discretized into a mixture of particles and a continuous-time gradient descent is performed on their weights and positions. This is an idealization of the usual way to train neural networks with a large hidden layer. We show that, when initialized correctly and in the many-particle limit, this gradient flow, although non-convex, converges to global minimizers. The proof involves Wasserstein gradient flows, a by-product of optimal transport theory. Numerical experiments show that this asymptotic behavior is already at play for a reasonable number of particles, even in high dimension.

## 7.2. Sharp Analysis of Learning with Discrete Losses

In [49], we study a least-squares framework to systematically design learning algorithms for discrete losses, with quantitative characterizations in terms of statistical and computational complexity. In particular we improve existing results by providing explicit dependence on the number of labels for a wide class of losses and faster learning rates in conditions of low-noise. Theoretical results are complemented with experiments on real datasets, showing the effectiveness of the proposed general approach.

## 7.3. Gossip of Statistical Observations using Orthogonal Polynomials

Consider a network of agents connected by communication links, where each agent holds a real value. The gossip problem consists in estimating the average of the values diffused in the network in a distributed manner. Current techniques for gossiping are designed to deal with worst-case scenarios, which is irrelevant in applications to distributed statistical learning and denoising in sensor networks. In [39], we design second-order gossip methods tailor-made for the case where the real values are i.i.d. samples from the same distribution. In some regular network structures, we are able to prove optimality of our methods, and simulations suggest that they are efficient in a wide range of random networks. Our approach of gossip stems from a new acceleration framework using the family of orthogonal polynomials with respect to the spectral measure of the network graph.

## 7.4. Marginal Weighted Maximum Log-likelihood for Efficient Learning of Perturb-and-Map models

In [20], We consider the structured-output prediction problem through probabilistic approaches and generalize the "perturb-and-MAP" framework to more challenging weighted Hamming losses, which are crucial in applications. While in principle our approach is a straightforward marginalization, it requires solving many related MAP inference problems. We show that for log-supermodular pairwise models these operations can be performed efficiently using the machinery of dynamic graph cuts. We also propose to use double stochastic gradient descent, both on the data and on the perturbations, for efficient learning. Our framework can naturally take weak supervision (e.g., partial labels) into account. We conduct a set of experiments on medium-scale character recognition and image segmentation, showing the benefits of our algorithms.

## 7.5. Slice inverse regression with score functions

In [6], we consider non-linear regression problems where we assume that the response depends non-linearly on a linear projection of the covariates. We propose score function extensions to sliced inverse regression problems, both for the first- order and second-order score functions. We show that they provably improve estimation in the population case over the non-sliced versions and we study finite sample estimators and their consistency given the exact score functions. We also propose to learn the score function as well, in two steps, i.e., first learning the score function and then learning the effective dimension reduction space, or directly, by solving a convex optimization problem regularized by the nuclear norm. We illustrate our results on a series of experiments.

## 7.6. Constant Step Size Stochastic Gradient Descent for Probabilistic Modeling

Stochastic gradient methods enable learning probabilistic models from large amounts of data. While large step-sizes (learning rates) have shown to be best for least-squares (e.g., Gaussian noise) once combined with parameter averaging, these are not leading to convergent algorithms in general. In this paper, we consider generalized linear models, that is, conditional models based on exponential families. In [14], we propose averaging moment parameters instead of natural parameters for constant-step-size stochastic gradient descent. For finite-dimensional models, we show that this can sometimes (and surprisingly) lead to better predictions than the best linear model. For infinite-dimensional models, we show that it always converges to optimal predictions, while averaging natural parameters never does. We illustrate our findings with simulations on synthetic data and classical benchmarks with many observations.

## 7.7. Nonlinear Acceleration of Momentum and Primal-Dual Algorithms

In [40], We describe a convergence acceleration scheme for multistep optimization algorithms. The extrapolated solution is written as a nonlinear average of the iterates produced by the original optimization algorithm. Our scheme does not need the underlying fixed-point operator to be symmetric, hence handles e.g. algorithms with momentum terms such as Nesterov's accelerated method, or primal-dual methods. The weights are computed via a simple linear system and we analyze performance in both online and offline modes. We use Crouzeix's conjecture to show that acceleration performance is controlled by the solution of a Chebyshev problem on the numerical range of a non-symmetric operator modelling the behavior of iterates near the optimum. Numerical experiments are detailed on image processing problems, logistic regression and neural network training for CIFAR10 and ImageNet.

## 7.8. Nonlinear Acceleration of Deep Neural Networks

Regularized nonlinear acceleration (RNA) is a generic extrapolation scheme for optimization methods, with marginal computational overhead. It aims to improve convergence using only the iterates of simple iterative algorithms. However, so far its application to optimization was theoretically limited to gradient descent and other single-step algorithms. Here, we adapt RNA to a much broader setting including stochastic gradient with momentum and Nesterov's fast gradient. In [54], we use it to train deep neural networks, and empirically observe that extrapolated networks are more accurate, especially in the early iterations. A straightforward application of our algorithm when training ResNet-152 on ImageNet produces a top-1 test error of 20.88, improving by 0.8 the reference classification pipeline. Furthermore, the code runs offline in this case, so it never negatively affects performance.

## 7.9. Nonlinear Acceleration of CNNs

The Regularized Nonlinear Acceleration (RNA) algorithm is an acceleration method capable of improving the rate of convergence of many optimization schemes such as gradient descend, SAGA or SVRG. Until now, its analysis is limited to convex problems, but empirical observations shows that RNA may be extended to wider settings. In [36], we investigate further the benefits of RNA when applied to neural networks, in particular for the task of image recognition on CIFAR10 and ImageNet. With very few modifications of exiting frameworks, RNA improves slightly the optimization process of CNNs, after training.

## 7.10. Robust Seriation and Applications To Cancer Genomics

The seriation problem seeks to reorder a set of elements given pairwise similarity information, so that elements with higher similarity are closer in the resulting sequence. When a global ordering consistent with the similarity information exists, an exact spectral solution recovers it in the noiseless case and seriation is equivalent to the combinatorial 2-SUM problem over permutations, for which several relaxations have been derived. However, in applications such as DNA assembly, similarity values are often heavily corrupted, and the solution of 2-SUM may no longer yield an approximate serial structure on the elements. In [52], we introduce the robust seriation problem and show that it is equivalent to a modified 2-SUM problem for a class of similarity matrices modeling those observed in DNA assembly. We explore several relaxations of this modified 2-SUM problem and compare them empirically on both synthetic matrices and real DNA data. We then introduce the problem of seriation with duplications, which is a generalization of Seriation motivated by applications to cancer genome reconstruction. We propose an algorithm involving robust seriation to solve it, and present preliminary results on synthetic data sets.

## 7.11. Reconstructing Latent Orderings by Spectral Clustering

Spectral clustering uses a graph Laplacian spectral embedding to enhance the cluster structure of some data sets. When the embedding is one dimensional, it can be used to sort the items (spectral ordering). A number of empirical results also suggests that a multidimensional Laplacian embedding enhances the latent ordering of the data, if any. This also extends to circular orderings, a case where unidimensional embeddings fail. In [51], we tackle the task of retrieving linear and circular orderings in a unifying framework, and show how a latent ordering on the data translates into a filamentary structure on the Laplacian embedding. We propose a method to recover it, illustrated with numerical experiments on synthetic data and real DNA sequencing data.

## 7.12. Lyapunov Functions for First-Order Methods: Tight Automated Convergence Guarantees

In [21], we present a novel way of generating Lyapunov functions for proving linear convergence rates of first-order optimization methods. Our approach provably obtains the fastest linear convergence rate that can be verified by a quadratic Lyapunov function (with given states), and only relies on solving a small-sized semidefinite program. Our approach combines the advantages of performance estimation problems and integral quadratic constraints, and relies on convex interpolation.

## 7.13. Efficient First-order Methods for Convex Minimization: a Constructive Approach

In [44], we describe a novel constructive technique for devising efficient first-order methods for a wide range of large-scale convex minimization settings, including smooth, non-smooth, and strongly convex minimization. The design technique takes a method performing a series of subspace-searches and constructs a family of methods that share the same worst-case guarantees as the original method, and includes a fixed-step first-order method. We show that this technique yields optimal methods in the smooth and non-smooth cases and derive new methods for these cases, including methods that forego knowledge of the problem parameters, at the cost of a one-dimensional line search per iteration. In the strongly convex case, we show how numerical tools can be used to perform the construction, and show that resulting method offers an improved convergence rate compared to Nesterov's celebrated fast gradient method.

## 7.14. Operator Splitting Performance Estimation: Tight contraction factors and optimal parameter selection

In [53], we propose a methodology for studying the performance of common splitting methods through semidefinite programming. We prove tightness of the methodology and demonstrate its value by presenting

two applications of it. First, we use the methodology as a tool for computer-assisted proofs to prove tight analytical contraction factors for Douglas–Rachford splitting that are likely too complicated for a human to find bare-handed. Second, we use the methodology as an algorithmic tool to computationally select the optimal splitting method parameters by solving a series of semidefinite programs.

## 7.15. Finite-sample Analysis of M-estimators using Self-concordance

In [50], we demonstrate how *self-concordance* of the loss allows to obtain asymptotically optimal rates for $M$-estimators in finite-sample regimes. We consider two classes of losses: (i) self-concordant losses, i.e., whose third derivative is uniformly bounded with the $3/2$ power of the second; (ii) *pseudo* self-concordant losses, for which the power is removed. These classes contain some losses arising in generalized linear models, including the logistic loss; in addition, the second class includes some common pseudo-Huber losses. Our results consist in establishing the *critical sample size* sufficient to reach the asymptotically optimal excess risk in both cases. Denoting $d$ the parameter dimension, and $d_e$ the effective dimension taking into account possible model misspecification, we find the critical sample size to be $O(d_e \cdot d)$ for the first class of losses, and $O(\rho \cdot d_e \cdot d)$ for the second class, where $\rho$ is the problem-dependent parameter that characterizes the risk curvature at the best predictor $\theta_*$. In contrast to the existing results, we only impose *local* assumptions on the data distribution, assuming that the calibrated design, i.e., the design scaled with the square root of the second derivative of the loss, is subgaussian at the best predictor. Moreover, we obtain the improved bounds on the critical sample size, scaling *near-linearly* in $\max(d_e, d)$, under the extra assumption that the calibrated design is subgaussian in the Dikin ellipsoid of $\theta_*$. Motivated by these findings, we construct canonically self-concordant analogues of the Huber and logistic losses with improved statistical properties. Finally, we extend some of the above results to $\ell_1$-penalized $M$-estimators in high-dimensional setups.

## 7.16. Uniform regret bounds over $R^d$ for the sequential linear regression problem with the square loss

In [45] we consider the setting of online linear regression for arbitrary deterministic sequences, with the square loss. We are interested in obtaining regret bounds that hold uniformly over all vectors $R^d$. When the feature sequence is known at the beginning of the game, they provided closed-form regret bounds of $2dB^2 \ln T + O(1)$, where $T$ is the number of rounds and $B$ is a bound on the observations. Instead, we derive bounds with an optimal constant of 1 in front of the $dB^2 \ln T$ term. In the case of sequentially revealed features, we also derive an asymptotic regret bound of $dB^2 \ln T$ for any individual sequence of features and bounded observations. All our algorithms are variants of the online nonlinear ridge regression forecaster, either with a data-dependent regularization or with almost no regularization.

## 7.17. Efficient online algorithms for fast-rate regret bounds under sparsity.

In [46] we consider the problem of online convex optimization in two different settings: arbitrary and i.i.d. sequence of convex loss functions. In both settings, we provide efficient algorithms whose cumulative excess risks are controlled with fast-rate sparse bounds. First, the excess risks bounds depend on the sparsity of the objective rather than on the dimension of the parameters space. Second, their rates are faster than the slow-rate $1/\sqrt{T}$

## 7.18. Exponential convergence of testing error for stochastic gradient methods

In [32], we consider binary classification problems with positive definite kernels and square loss, and study the convergence rates of stochastic gradient methods. We show that while the excess testing loss (squared loss) converges slowly to zero as the number of observations (and thus iterations) goes to infinity, the testing error (classification error) converges exponentially fast if low-noise conditions are assumed.

## 7.19. Statistical Optimality of Stochastic Gradient Descent on Hard Learning Problems through Multiple Passes

In [33], we consider stochastic gradient descent (SGD) for least-squares regression with potentially several passes over the data. While several passes have been widely reported to perform practically better in terms of predictive performance on unseen data, the existing theoretical analysis of SGD suggests that a single pass is statistically optimal. While this is true for low-dimensional easy problems, we show that for hard problems, multiple passes lead to statistically optimal predictions while single pass does not; we also show that in these hard models, the optimal number of passes over the data increases with sample size. In order to define the notion of hardness and show that our predictive performances are optimal, we consider potentially infinite-dimensional models and notions typically associated to kernel methods, namely, the decay of eigenvalues of the covariance matrix of the features and the complexity of the optimal predictor as measured through the covariance matrix. We illustrate our results on synthetic experiments with non-linear kernel methods and on a classical benchmark with a linear model.

## 7.20. Central Limit Theorem for stationary Fleming–Viot particle systems in finite spaces

In [11], we consider the Fleming–Viot particle system associated with a continuous-time Markov chain in a finite space. Assuming irreducibility, it is known that the particle system possesses a unique stationary distribution, under which its empirical measure converges to the quasistationary distribution of the Markov chain. We complement this Law of Large Numbers with a Central Limit Theorem. Our proof essentially relies on elementary computations on the infinitesimal generator of the Fleming–Viot particle system, and involves the so-called $\pi$-return process in the expression of the asymptotic variance. Our work can be seen as an infinite-time version, in the setting of finite space Markov chains, of results by Del Moral and Miclo [ESAIM: Probab. Statist., 2003] and Cérou, Delyon, Guyader and Rousset [arXiv:1611.00515, arXiv:1709.06771].

## 7.21. SeaRNN: Improved RNN training through Global-Local Losses

In [16], we propose SEARNN, a novel training algorithm for recurrent neural networks (RNNs) inspired by the "learning to search" (L2S) approach to structured prediction. RNNs have been widely successful in structured prediction applications such as machine translation or parsing, and are commonly trained using maximum likelihood estimation (MLE). Unfortunately, this training loss is not always an appropriate surrogate for the test error: by only maximizing the ground truth probability, it fails to exploit the wealth of information offered by structured losses. Further, it introduces discrepancies between training and predicting (such as exposure bias) that may hurt test performance. Instead, SEARNN leverages test-alike search space exploration to introduce global-local losses that are closer to the test error. We first demonstrate improved performance over MLE on two different tasks: OCR and spelling correction. Then, we propose a subsampling strategy to enable SEARNN to scale to large vocabulary sizes. This allows us to validate the benefits of our approach on a machine translation task.

## 7.22. Improved asynchronous parallel optimization analysis for stochastic incremental methods

As datasets continue to increase in size and multi-core computer architectures are developed, asynchronous parallel optimization algorithms become more and more essential to the field of Machine Learning. Unfortunately, conducting the theoretical analysis of asynchronous methods is difficult, notably due to the introduction of delay and inconsistency in inherently sequential algorithms. Handling these issues often requires resorting to simplifying but unrealistic assumptions. Through a novel perspective, in [10] we revisit and clarify a subtle but important technical issue present in a large fraction of the recent convergence rate proofs for asynchronous parallel optimization algorithms, and propose a simplification of the recently introduced "perturbed iterate" framework that resolves it. We demonstrate the usefulness of our new framework by analyzing three distinct

asynchronous parallel incremental optimization algorithms: Hogwild (asynchronous SGD), KROMAGNON (asynchronous SVRG) and ASAGA, a novel asynchronous parallel version of the incremental gradient algorithm SAGA that enjoys fast linear convergence rates. We are able to both remove problematic assumptions and obtain better theoretical results. Notably, we prove that ASAGA and KROMAGNON can obtain a theoretical linear speedup on multi-core systems even without sparsity assumptions. We present results of an implementation on a 40-core architecture illustrating the practical speedups as well as the hardware overhead. Finally, we investigate the overlap constant, an ill-understood but central quantity for the theoretical analysis of asynchronous parallel algorithms. We find that it encompasses much more complexity than suggested in previous work, and often is order-of-magnitude bigger than traditionally thought.

## 7.23. Asynchronous optimisation for Machine Learning

The impressive breakthroughs of the last two decades in the field of machine learning can be in large part attributed to the explosion of computing power and available data. These two limiting factors have been replaced by a new bottleneck: algorithms. The focus of this thesis [3] is thus on introducing novel methods that can take advantage of high data quantity and computing power. We present two independent contributions.

First, we develop and analyze novel fast optimization algorithms which take advantage of the advances in parallel computing architecture and can handle vast amounts of data. We introduce a new framework of analysis for asynchronous parallel incremental algorithms, which enable correct and simple proofs. We then demonstrate its usefulness by performing the convergence analysis for several methods, including two novel algorithms.

Asaga is a sparse asynchronous parallel variant of the variance-reduced algorithm Saga which enjoys fast linear convergence rates on smooth and strongly convex objectives. We prove that it can be linearly faster than its sequential counterpart, even without sparsity assump- tions.

ProxAsaga is an extension of Asaga to the more general setting where the regularizer can be non-smooth. We prove that it can also achieve a linear speedup. We provide extensive experiments comparing our new algorithms to the current state-of-art.

Second, we introduce new methods for complex struc- tured prediction tasks. We focus on recurrent neural net- works (RNNs), whose traditional training algorithm for RNNs – based on maximum likelihood estimation (MLE) – suffers from several issues. The associated surrogate training loss notably ignores the information contained in structured losses and introduces discrepancies between train and test times that may hurt performance.

To alleviate these problems, we propose SeaRNN, a novel training algorithm for RNNs inspired by the "learning to search" approach to structured prediction.SeaRNN leverages test-alike search space exploration to introduce global-local losses that are closer to the test error than the MLE objective.

We demonstrate improved performance over MLE on three challenging tasks, and provide several subsampling strategies to enable SeaRNN to scale to large-scale tasks, such as machine translation. Finally, after contrasting the behavior of SeaRNN models to MLE models, we conduct an in-depth comparison of our new approach to the related work.

## 7.24. $M^*$-Regularized Dictionary Learning

In [38], we derive a performance measure for dictionaries in compressed sensing, based on the $M^*$ of the corresponding norm. We use this measure to regularize dictionary learning algorithms and study the performance of our methods on both compression and inpainting experiments.

## 7.25. Optimal Algorithms for Non-Smooth Distributed Optimization in Networks

In [35], we consider the distributed optimization of non-smooth convex functions using a network of computing units. We investigate this problem under two regularity assumptions: (1) the Lipschitz continuity

of the global objective function, and (2) the Lipschitz continuity of local individual functions. Under the local regularity assumption, we provide the first optimal first-order decentralized algorithm called multi-step primal-dual (MSPD) and its corresponding optimal convergence rate. A notable aspect of this result is that, for non-smooth functions, while the dominant term of the error is in $O(1/\sqrt{t})$, the structure of the communication network only impacts a second-order term in $O(1/t)$, where $t$ is time. In other words, the error due to limits in communication resources decreases at a fast rate even in the case of non-strongly-convex objective functions. Under the global regularity assumption, we provide a simple yet efficient algorithm called distributed randomized smoothing (DRS) based on a local smoothing of the objective function, and show that DRS is within a $d^{1/4}$ multiplicative factor of the optimal convergence rate, where $d$ is the underlying dimension.

## 7.26. Relating Leverage Scores and Density using Regularized Christoffel Functions

Statistical leverage scores emerged as a fundamental tool for matrix sketching and column sampling with applications to low rank approximation, regression, random feature learning and quadrature. Yet, the very nature of this quantity is barely understood. Borrowing ideas from the orthogonal polynomial literature, we introduce in [31] the regularized Christoffel function associated to a positive definite kernel. This uncovers a variational formulation for leverage scores for kernel methods and allows to elucidate their relationships with the chosen kernel as well as population density. Our main result quantitatively describes a decreasing relation between leverage score and population density for a broad class of kernels on Euclidean spaces. Numerical simulations support our findings.

## 7.27. Averaging Stochastic Gradient Descent on Riemannian Manifolds

In [37] we consider the minimization of a function defined on a Riemannian manifold $M$ accessible only through unbiased estimates of its gradients. We develop a geometric framework to transform a sequence of slowly converging iterates generated from stochastic gradient descent (SGD) on $M$ to an averaged iterate sequence with a robust and fast $O(1/n)$ convergence rate. We then present an application of our framework to geodesically-strongly-convex (and possibly Euclidean non-convex) problems. Finally, we demonstrate how these ideas apply to the case of streaming $k$-PCA, where we show how to accelerate the slow rate of the randomized power method (without requiring knowledge of the eigengap) into a robust algorithm achieving the optimal rate of convergence.

## 7.28. Localized Structured Prediction

Key to structured prediction is exploiting the problem structure to simplify the learning process. A major challenge arises when data exhibit a local structure (e.g., are made by "parts") that can be leveraged to better approximate the relation between (parts of) the input and (parts of) the output. Recent literature on signal processing, and in particular computer vision, has shown that capturing these aspects is indeed essential to achieve state-of-the-art performance. While such algorithms are typically derived on a case-by-case basis, in [42] we propose the first theoretical framework to deal with part-based data from a general perspective. We derive a novel approach to deal with these problems and study its generalization properties within the setting of statistical learning theory. Our analysis is novel in that it explicitly quantifies the benefits of leveraging the part-based structure of the problem with respect to the learning rates of the proposed estimator.

## 7.29. Optimal rates for spectral algorithms with least-squares regression over Hilbert spaces

In [12], we study regression problems over a separable Hilbert space with the square loss, covering non-parametric regression over a reproducing kernel Hilbert space. We investigate a class of spectral-regularized algorithms, including ridge regression, principal component analysis, and gradient methods. We prove optimal,

high-probability convergence results in terms of variants of norms for the studied algorithms, considering a capacity assumption on the hypothesis space and a general source condition on the target function. Consequently, we obtain almost sure convergence results with optimal rates. Our results improve and generalize previous results, filling a theoretical gap for the non-attainable cases.

## 7.30. Differential Properties of Sinkhorn Approximation for Learning with Wasserstein Distance

Applications of optimal transport have recently gained remarkable attention thanks to the computational advantages of entropic regularization. However, in most situations the Sinkhorn approximation of the Wasserstein distance is replaced by a regularized version that is less accurate but easy to differentiate. In [17] we characterize the differential properties of the original Sinkhorn distance, proving that it enjoys the same smoothness as its regularized version and we explicitly provide an efficient algorithm to compute its gradient. We show that this result benefits both theory and applications: on one hand, high order smoothness confers statistical guarantees to learning with Wasserstein approximations. On the other hand, the gradient formula allows us to efficiently solve learning and optimization problems in practice. Promising preliminary experiments complement our analysis.

## 7.31. Learning with SGD and Random Features

Sketching and stochastic gradient methods are arguably the most common techniques to derive efficient large scale learning algorithms. In [15], we investigate their application in the context of nonparametric statistical learning. More precisely, we study the estimator defined by stochastic gradient with mini batches and random features. The latter can be seen as form of nonlinear sketching and used to define approximate kernel methods. The considered estimator is not explicitly penalized/constrained and regularization is implicit. Indeed, our study highlights how different parameters, such as number of features, iterations, step-size and mini-batch size control the learning properties of the solutions. We do this by deriving optimal finite sample bounds, under standard assumptions. The obtained results are corroborated and illustrated by numerical experiments.

## 7.32. Manifold Structured Prediction

Structured prediction provides a general framework to deal with supervised problems where the outputs have semantically rich structure. While classical approaches consider finite, albeit potentially huge, output spaces, in [19] we discuss how structured prediction can be extended to a continuous scenario. Specifically, we study a structured prediction approach to manifold valued regression. We characterize a class of problems for which the considered approach is statistically consistent and study how geometric optimization can be used to compute the corresponding estimator. Promising experimental results on both simulated and real data complete our study.

## 7.33. On Fast Leverage Score Sampling and Optimal Learning

Leverage score sampling provides an appealing way to perform approximate computations for large matrices. Indeed, it allows to derive faithful approximations with a complexity adapted to the problem at hand. Yet, performing leverage scores sampling is a challenge in its own right requiring further approximations. In [18], we study the problem of leverage score sampling for positive definite matrices defined by a kernel. Our contribution is twofold. First we provide a novel algorithm for leverage score sampling and second, we exploit the proposed method in statistical learning by deriving a novel solver for kernel ridge regression. Our main technical contribution is showing that the proposed algorithms are currently the most efficient and accurate for these problems.

### 7.34. Accelerated Decentralized Optimization with Local Updates for Smooth and Strongly Convex Objectives

In [47], we study the problem of minimizing a sum of smooth and strongly convex functions split over the nodes of a network in a decentralized fashion. We propose a decentralized accelerated algorithm that only requires local synchrony. Its rate depends on the condition number $\kappa$ of the local functions as well as the network topology and delays. Under mild assumptions on the topology of the graph, our algorithm takes a time $O((\tau_{\max} + \Delta_{\max})\sqrt{\kappa/\gamma}\ln(\epsilon^{-1}))$ to reach a precision $\epsilon$ where $\gamma$ is the spectral gap of the graph, $\tau_{\max}$ the maximum communication delay and $\Delta_{\max}$ the maximum computation time. Therefore, it matches the rate of SSDA, which is optimal when $\tau_{\max} = \Omega(\Delta_{\max})$. Applying our algorithm to quadratic local functions leads to an accelerated randomized gossip algorithm of rate $O(\sqrt{\theta_{\text{gossip}}/n})$ where $\theta_{\text{gossip}}$ is the rate of the standard randomized gossip. To the best of our knowledge, it is the first asynchronous gossip algorithm with a provably improved rate of convergence of the second moment of the error. We illustrate these results with experiments in idealized settings.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

Microsoft Research: “Structured Large-Scale Machine Learning”. Machine learning is now ubiquitous in industry, science, engineering, and personal life. While early successes were obtained by applying off-the-shelf techniques, there are two main challenges faced by machine learning in the “big data” era: structure and scale. The project proposes to explore three axes, from theoretical, algorithmic and practical perspectives: (1) large-scale convex optimization, (2) large-scale combinatorial optimization and (3) sequential decision making for structured data. The project involves two Inria sites (Paris and Grenoble) and four MSR sites (Cambridge, New England, Redmond, New York). Project website: http://www.msr-inria.fr/projects/structured-large-scale-machine-learning/.

## 8.2. Bilateral Grants with Industry

- Alexandre d’Aspremont, Francis Bach, Martin Jaggi (EPFL): Google Focused award.
- Francis Bach: Gift from Facebook AI Research.
- Alexandre d’Aspremont: AXA, "mécénat scientifique, chaire Havas-Dauphine", machine learning.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

Alexandre d’Aspremont: IRIS, PSL “Science des données, données de la science”.

## 9.2. European Initiatives

- **ITN Spartan**
  Title: Sparse Representations and Compressed Sensing Training Network
  Type: FP7
  Instrument: Initial Training Network
  Duration: October 2014 to October 2018
  Coordinator: Mark Plumbley (University of Surrey)
  Inria contact: Francis Bach
  Abstract: The SpaRTaN Initial Training Network will train a new generation of interdisciplinary researchers in sparse representations and compressed sensing, contributing to Europe’s leading role

in scientific innovation. By bringing together leading academic and industry groups with expertise in sparse representations, compressed sensing, machine learning and optimisation, and with an interest in applications such as hyperspectral imaging, audio signal processing and video analytics, this project will create an interdisciplinary, trans-national and inter-sectorial training network to enhance mobility and training of researchers in this area. SpaRTaN is funded under the FP7-PEOPLE-2013-ITN call and is part of the Marie Curie Actions — Initial Training Networks (ITN) funding scheme: Project number - 607290

- **ITN Macsenet**
  Title: Machine Sensing Training Network
  Type: H2020
  Instrument: Initial Training Network
  Duration: January 2015 - January 2019
  Coordinator: Mark Plumbley (University of Surrey)
  Inria contact: Francis Bach
  Abstract: The aim of this Innovative Training Network is to train a new generation of creative, entrepreneurial and innovative early stage researchers (ESRs) in the research area of measurement and estimation of signals using knowledge or data about the underlying structure. We will develop new robust and efficient Machine Sensing theory and algorithms, together methods for a wide range of signals, including: advanced brain imaging; inverse imaging problems; audio and music signals; and non-traditional signals such as signals on graphs. We will apply these methods to real-world problems, through work with non-Academic partners, and disseminate the results of this research to a wide range of academic and non-academic audiences, including through publications, data, software and public engagement events. MacSeNet is funded under the H2020-MSCA-ITN-2014 call and is part of the Marie Sklodowska- Curie Actions — Innovative Training Networks (ITN) funding scheme.
- **ERC Sequoia** Title: Robust algorithms for learning from modern data
  Programm: H2020
  Type: ERC
  Duration: 2017-2022
  Coordinator: Inria
  Inria contact: Francis Bach
  Abstract: Machine learning is needed and used everywhere, from science to industry, with a growing impact on many disciplines. While first successes were due at least in part to simple supervised learning algorithms used primarily as black boxes on medium-scale problems, modern data pose new challenges. Scalability is an important issue of course: with large amounts of data, many current problems far exceed the capabilities of existing algorithms despite sophisticated computing architectures. But beyond this, the core classical model of supervised machine learning, with the usual assumptions of independent and identically distributed data, or well-defined features, outputs and loss functions, has reached its theoretical and practical limits. Given this new setting, existing optimization-based algorithms are not adapted. The main objective of this project is to push the frontiers of supervised machine learning, in terms of (a) scalability to data with massive numbers of observations, features, and tasks, (b) adaptability to modern computing environments, in particular for parallel and distributed processing, (c) provable adaptivity and robustness to problem and hardware specifications, and (d) robustness to non-convexities inherent in machine learning problems. To achieve the expected breakthroughs, we will design a novel generation of learning algorithms amenable to a tight convergence analysis with realistic assumptions and efficient implementations. They will help transition machine learning algorithms towards the same widespread robust use as numerical linear algebra libraries. Outcomes of the research described in this proposal will include algorithms that come with strong convergence guarantees and are well-tested on real-life benchmarks coming from computer vision, bioin- formatics, audio processing and natural

language processing. For both distributed and non-distributed settings, we will release open-source software, adapted to widely available computing platforms.

## 9.3. International Initiatives

### *9.3.1. BigFOKS2*

Title: Learning from Big Data: First-Order methods for Kernels and Submodular functions

International Partner (Institution - Laboratory - Researcher):

IISc Bangalore (India) - Computer Science Department - Chiranjib Bhattacharyya

Start year: 2016

See also: mllab.csa.iisc.ernet.in/indo-french.html

Recent advances in sensor technologies have resulted in large amounts of data being generated in a wide array of scientific disciplines. Deriving models from such large datasets, often known as "Big Data", is one of the important challenges facing many engineering and scientific disciplines. In this proposal we investigate the problem of learning supervised models from Big Data, which has immediate applications in Computational Biology, Computer vision, Natural language processing, Web, E-commerce, etc., where specific structure is often present and hard to take into account with current algorithms. Our focus will be on the algorithmic aspects. Often supervised learning problems can be cast as convex programs. The goal of this proposal will be to derive first-order methods which can be effective for solving such convex programs arising in the Big-Data setting. Keeping this broad goal in mind we investigate two foundational problems which are not well addressed in existing literature. The first problem investigates Stochastic Gradient Descent Algorithms in the context of First-order methods for designing algorithms for Kernel based prediction functions on Large Datasets. The second problem involves solving discrete optimization problems arising in Submodular formulations in Machine Learning, for which first-order methods have not reached the level of speed required for practical applications (notably in computer vision).

## 9.4. International Research Visitors

- Vijaya Bollapragada from Northwestern University, Chicago, IL, United States, Apr - Jul 2018.
- Aaron De Fazio from Facebook Research NY, New York, United States, Feb 2018.
- Gauthier Gidel from University of Montreal - MILA, Montreal, Canada, Jan 2018.
- Sharan Vaswani from University of British Columbia, Vancouver, Canada, Apr - Jul 2018
- Simon Lacoste-Julien from University of Montreal - MILA, Montreal, Canada, Aug 2018.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

F. Bach: General Chair of ICML 2018 (Stockholm)

*10.1.1.2. Member of the Organizing Committees*

Adrian Taylor, Session Organizer: *Computer-assisted analyses of optimization algorithms I & II*, International Symposium on Mathematical Programming, July 2018.

F. Bach: Co-organization of the workshop "Horizon Maths 2018 : Intelligence Artificielle", November 23, 2018

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Chair of Conference Program Committees*

F. Bach: Program Chair of the Journées de Statistiques (Saclay)

*10.1.2.2. Reviewer*

Conference on Learning Theory (COLT 2018): Pierre Gaillard, Alessandro Rudi

Symposium on Discrete Algorithms (SODA 2019): Adrien Taylor,

Neural Information Processing Systems (NIPS 2018): Pierre Gaillard, Alessandro Rudi

Conference on Learning Theory (COLT 2018): Pierre Gaillard, Alessandro Rudi, Adrien Taylor

Symposium on Discrete Algorithms (SODA 2019): Adrien Taylor

International Conference of Machine Learning (ICML 2018): Pierre Gaillard, Alessandro Rudi

### *10.1.3. Journal*

*10.1.3.1. Member of the Editorial Boards*

F. Bach: Journal of Machine Learning Research, co-editor-in-chief

F. Bach: Information and Inference, Associate Editor.

F. Bach: Electronic Journal of Statistics, Associate Editor.

F. Bach: Mathematical Programming, Associate Editor.

F. Bach: Foundations of Computational Mathematics, Associate Editor.

A. d'Aspremont: SIAM Journal on Optimization, Associate editor

A. d'Aspremont: SIAM Journal on the Mathematics of Data Science, Associate Editor

A. d'Aspremont: Mathematical Programming, Associate Editor

*10.1.3.2. Reviewer - Reviewing Activities*

SIAM Journal on Optimization: Adrien Taylor

Mathematical Programming: Adrien Taylor

Journal of Optimization Theory and Algorithms: Adrien Taylor

Journal of Machine Learning Research: Pierre Gaillard, Alessandro Rudi

Applied Computational Harmonic Analysis: Alessandro Rudi

### *10.1.4. Invited Talks*

F. Bach, Trends in Optimization Seminar, University of Washington, November 2018.

Pierre Gaillard. *Distributed averaging of observations in a graph: the gossip problem*. MNL Conference, Paris, November 2018.

Adrien Taylor, *Analysis and design of first-order methods via semidefinite programming*, Seminaire Parisien dOptimisation (SPO), Paris (France), November 2018.

F. Bach, Frontier Research and Artificial Intelligence, European Research Council, Brussels, October 2018.

F. Bach, IDSS Distinguished Speaker Seminar, MIT, October 2018.

F. Bach, Mathematical Institute Colloquium, Oxford, October 2018.

Adrien Taylor, *Convex Interpolation and Performance Estimation of First- order Methods* for Convex Optimization, IBM/FNRS innovation award, Brussels (Belgium), October 2018.

F. Bach, Workshop on Structural Inference in High-Dimensional Models, Moscow, September 2018.

F. Bach, Symposium on Mathematical Programming (ISMP), Bordeaux, plenary talk, July 2018.

Alexandre d'Aspremont, *Sharpness, Restart and Compressed Sensing Performance*, ISMP 2018, Bordeaux, July 2018.

Alessandro Rudi, *FALKON: An optimal method for large scale learning with statistical guarantees*, ISMP 2018, Bordeaux, July 2018.

Adrien Taylor, *Computer-assisted Lyapunov-based worst-case analyses of first- order methods*, International Symposium on Mathematical Programming, Bordeaux (France), July 2018.

F. Bach, SIAM Conference on Imaging Science, Bologna, Italy, invited talk, June 2018.

Pierre Gaillard. *Online prediction of arbitrary time-series with application to electricity consumption*. Conference on nonstationarity. Cergy Pontoise University. June 2018.

Adrien Taylor, *Convex Interpolation and Performance Estimation of First-order Methods for Convex Optimization*, International Symposium on Mathematical Programming: Tucker prize finalist, Bordeaux (France), July 2018.

Alexandre d'Aspremont, *An approximate Shapley-Folkman Theorem*, Isaac Newton Institute, Cambridge, June 2018.

F. Bach,Workshop on Future challenges in statistical scalability, Newton Institute, Cambridge, UK, June 2018.

Adrien Taylor, *Automated design of first-order optimization methods*, Operation Research Seminar, UCLouvain, Louvain-la-Neuve (Belgium), May 2018.

Adrien Taylor, *Automated design of first-order optimization methods*, LCCC Control Seminar, Lund University, Lund (Sweden), May 2018.

Pierre Gaillard. *Distributed learning with orthogonal polynomials*. Inria DGA meetup. May 2018.

F. Bach, Workshop on Optimisation and Machine Learning in Economics, London, March 2018.

Pierre Gaillard. *An overview of Artificial Intelligence*. Hackaton. PSL University. March 2018.

Alexandre d'Aspremont, *Regularized Nonlinear Acceleration*, US and Mexico Workshop on Optimization and its Applications, Jan 2018.

Alessandro Rudi, *Learning with Random Features*, Isaac Newton Institute, Cambridge, Jan 2018.

Pierre Gaillard. *Online nonparametric regression with adversarial data.* Smile seminar. Paris. Jan 2018.

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

F. Bach (together with N. Chopin), *Graphical models*, 30h, Master M2 (MVA), ENS Cachan, France.

F. Bach, *Optimisation et apprentissage statistique*, 20h, Master M2 (Mathématiques de l'aléatoire), Université Paris-Sud, France.

Alexandre d'Aspremont, *Optimisation Combinatoire et Convexe*, avec Zhentao Li, (2015-Present) cours magistraux 30h, Master M1, ENS Paris.

Alexandre d'Aspremont, *Optimisation convexe: modélisation, algorithmes et applications* cours magistraux 21h (2011-Present), Master M2 MVA, ENS PS.

F. Bach and P. Gaillard, *Apprentissage statistique*, 35h, Master M1, Ecole Normale Supérieure, France.

P. Gaillard (together with V. Perchet), *Prediction of individual sequences*, 21h, Master M2 MVA, ENS Cachan, France.

Gregoire Mialon, Python for Machine Learning, 21h, M2 MASH, Dauphine-ENS-PSL, Paris.

### *10.2.2. Supervision*

Anaël Bonneton, PhD defended on July 2018, co-advised by Francis Bach, located in Agence nationale de la sécurité des systèmes d'information (ANSSI).

Damien Scieur, PhD defended on September 2018. *Sur l'accélération des méthodes d'optimisation*, supervised by Alexandre d'Aspremont and Francis Bach.

Jean-Baptiste Alayrac, PhD defended on September 2018, *Structured Learning from Videos and Language*, supervised by Simon Lacoste-Julien, Josef Sivic and Ivan Laptev.

Antoine Recanati, PhD. defended on November 2018. *Application du problème de sériation au séquençage de l'ADN et autres relaxations convexes appliquées en bioinformatique*, supervised by Alexandre d'Aspremont.

Rémi Leblond, PhD defended on November 2018, *Asynchronous Optimization for Machine Learning*, supervised by Simon Lacoste-Julien.

Mathieu Barre, PhD in progress *Méthodes d'extrapolation, au-delà de la convexité*, supervised by Alexandre d'Aspremont.

Grégoire Mialon, PhD in progress *Algorithmes d'optimisation, méthodes de régularisation et architectures pour les réseaux de neurones profonds dans un contexte où les données labellisées sont rares*, supervised by Alexandre d'Aspremont.

Radu-Alexandru Dragomir, PhD in progress *Non-Euclidean first-order methods*, supervised by Alexandre d'Aspremont and Jérôme Bolte.

Thomas Kerdreux, PhD in progress *Optimisation and machine learning*, supervised by Alexandre d'Aspremont.

Margaux Brégère, PhD in progress started September 2017, supervised by Pierre Gaillard, Gilles Stoltz and Yannig Goude (EDF R&D).

Raphaël Berthier, PhD in progress started September 2017, supervised by Francis Bach and Pierre Gaillard.

Loucas Pillaud-Vivien, PhD in progress, supervised by Francis Bach and Alessandro Rudi.

Alex Nowak, PhD in progress, supervised by Francis Bach and Alessandro Rudi.

Ulysse Marteau Ferey, PhD in progress, supervised by Francis Bach and Alessandro Rudi.

Dmitry Babichev, PhD in progress, started is September 2015, co-advised by Francis Bach and Anatoly Judistky (Univ. Grenoble).

Tatiana Shpakova, PhD in progress, started September 2015, advised by Francis Bach.

### *10.2.3. Juries*

Alexandre d'Aspremont, Habilitation à diriger des recherches. Thomas Bruls, Genoscope, Université d'Evry.

## 10.3. Popularization

### *10.3.1. Creation of media or tools for science outreach*

Design and implementation of a demonstration for the permanent exhibit at Palais de la Découverte: "L'apprenti illustrateur" (J.-B. Alayrac, F. Bach)

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] J.-B. ALAYRAC.*Structured Learning from Videos and Language*, Ecole normale supérieure - ENS PARIS, September 2018, https://hal.inria.fr/tel-01885412

[2] A. BEAUGNON.*Expert-in-the-Loop Supervised Learning for Computer Security Detection Systems*, PSL Research University, June 2018, https://hal.archives-ouvertes.fr/tel-01888971

[3] R. LEBLOND.*Asynchronous Optimization for Machine Learning*, Ecole Normale Superieure de Paris - ENS Paris, November 2018, https://hal.inria.fr/tel-01950576

[4] A. RECANATI.*Relaxations of the Seriation problem and applications to de novo genome assembly*, PSL Research University, November 2018, https://hal.archives-ouvertes.fr/tel-01984368

[5] D. SCIEUR.*Acceleration in Optimization*, PSL Research University, September 2018, https://hal.archives-ouvertes.fr/tel-01887163

### Articles in International Peer-Reviewed Journal

[6] D. BABICHEV, F. BACH.*Slice inverse regression with score functions*, in "Electronic journal of statistics ", May 2018, vol. Volume 12, Number 1 (2018), p. 1507-1543 [*DOI :* 10.1214/18-EJS1428], https://hal.inria.fr/hal-01388498

[7] F. BACH.*Submodular Functions: from Discrete to Continous Domains*, in "Mathematical Programming, Series A", 2018, https://arxiv.org/abs/1511.00394 , https://hal.archives-ouvertes.fr/hal-01222319

[8] A. D'ASPREMONT, C. GUZMAN, M. JAGGI.*Optimal Affine-Invariant Smooth Minimization Algorithms*, in "SIAM Journal on Optimization", July 2018, vol. 28, n$^{o}$ 3, p. 2384 - 2405 [*DOI :* 10.1137/17M1116842], https://hal.archives-ouvertes.fr/hal-01927392

[9] D. GARREAU, S. ARLOT.*Consistent change-point detection with kernels*, in "Electronic journal of statistics ", December 2018, vol. 12, n$^{o}$ 2, p. 4440-4486, https://arxiv.org/abs/1612.04740 , https://hal.archives-ouvertes.fr/hal-01416704

[10] R. LEBLOND, F. PEDREGOSA, S. LACOSTE-JULIEN.*Improved asynchronous parallel optimization analysis for stochastic incremental methods*, in "Journal of Machine Learning Research (JMLR)", 2018, https://hal.inria.fr/hal-01950558

[11] T. LELIEVRE, L. PILLAUD-VIVIEN, J. REYGNER.*Central Limit Theorem for stationary Fleming–Viot particle systems in finite spaces*, in "ALEA : Latin American Journal of Probability and Mathematical Statistics", September 2018, vol. 15, p. 1163-1182, https://arxiv.org/abs/1806.04490 [*DOI :* 10.30757/ALEA.v15-43], https://hal-enpc.archives-ouvertes.fr/hal-01812120

[12] J. LIN, A. RUDI, L. ROSASCO, V. CEVHER.*Optimal rates for spectral algorithms with least-squares regression over Hilbert spaces*, in "Applied and Computational Harmonic Analysis", October 2018, https://hal.inria.fr/hal-01958890

[13] T. SCHATZ, F. BACH, E. DUPOUX.*Evaluating automatic speech recognition systems as quantitative models of cross-lingual phonetic category perception*, in "Journal of the Acoustical Society of America", May 2018, vol. 143, n$^{o}$ 5, p. EL372 - EL378 [*DOI :* 10.1121/1.5037615], https://hal.archives-ouvertes.fr/hal-01888735

### International Conferences with Proceedings

[14] D. BABICHEV, F. BACH.*Constant Step Size Stochastic Gradient Descent for Probabilistic Modeling*, in "UAI 2018 - Conference on Uncertainty in Artificial Intelligence", Monterey, United States, August 2018, https://arxiv.org/abs/1804.05567 , https://hal.inria.fr/hal-01929810

[15] L. CARRATINO, A. RUDI, L. ROSASCO.*Learning with SGD and Random Features*, in "Advances in Neural Information Processing Systems", Montreal, Canada, December 2018, p. 10213–10224, https://arxiv.org/abs/1807.06343 - Spotlight, https://hal.archives-ouvertes.fr/hal-01958906

[16] R. LEBLOND, J.-B. ALAYRAC, A. OSOKIN, S. LACOSTE-JULIEN.*SeaRNN: Training RNNs with Global-Local Losses*, in "ICLR 2018 : 6th International Conference on Learning Representations", Vancouver, Canada, April 2018, https://hal.inria.fr/hal-01950555

[17] G. LUISE, A. RUDI, M. PONTIL, C. CILIBERTO.*Differential Properties of Sinkhorn Approximation for Learning with Wasserstein Distance*, in "NIPS 2018 - Advances in Neural Information Processing Systems", Montreal, Canada, December 2018, p. 5864-5874, https://arxiv.org/abs/1805.11897 - 26 pages, 4 figures, https://hal.inria.fr/hal-01958887

[18] A. RUDI, D. CALANDRIELLO, L. CARRATINO, L. ROSASCO.*On Fast Leverage Score Sampling and Optimal Learning*, in "NeurIPS 2018 - Thirty-second Conference on Neural Information Processing Systems", Montreal, Canada, Advances in Neural Information Processing Systems - NIPS-2018, December 2018, vol. 31, p. 5677–5687, https://arxiv.org/abs/1810.13258 , https://hal.inria.fr/hal-01958879

[19] A. RUDI, C. CILIBERTO, G. M. MARCONI, L. ROSASCO.*Manifold Structured Prediction*, in "NIPS 2018 - Neural Information Processing Systems Conference", Montreal, Canada, Advances in Neural Information Processing Systems, December 2018, vol. 31, p. 5615-5626, https://arxiv.org/abs/1806.09908 , https://hal.archives-ouvertes.fr/hal-01958900

[20] T. SHPAKOVA, F. BACH, A. OSOKIN.*Marginal Weighted Maximum Log-likelihood for Efficient Learning of Perturb-and-Map models*, in "UAI 2018 - Conference on Uncertainty in Artificial Intelligence 2018", Monterey, United States, August 2018, https://arxiv.org/abs/1811.08725 , https://hal.inria.fr/hal-01939549

[21] A. B. TAYLOR, B. VAN SCOY, L. LESSARD.*Lyapunov Functions for First-Order Methods: Tight Automated Convergence Guarantees*, in "Proceedings of the 35th International Conference on Machine Learning. PMLR 80:4897-4906", Stockholm, Sweden, July 2018, https://arxiv.org/abs/1803.06073 , https://hal.inria.fr/hal-01902068

### Conferences without Proceedings

[22] F. BACH.*Efficient Algorithms for Non-convex Isotonic Regression through Submodular Optimization*, in "Advances in Neural Information Processing Systems", Montreal, Canada, December 2018, https://arxiv.org/abs/1707.09157 , https://hal.archives-ouvertes.fr/hal-01569934

[23] A. BEAUGNON, P. CHIFFLIER, F. BACH.*End-to-End Active Learning for Computer Security Experts*, in "KDD Workshop on Interactive Data Exploration and Analytics (IDEA)", Londres, United Kingdom, August 2018, https://hal.archives-ouvertes.fr/hal-01888983

[24] A. BEAUGNON, P. CHIFFLIER, F. BACH.*End-to-End Active Learning for Computer Security Experts*, in "AAAI Workshop on Artificial Intelligence for Cyber Security (AICS)", New Orleans, United States, February 2018, https://hal.archives-ouvertes.fr/hal-01888976

[25] L. CHIZAT, F. BACH.*On the Global Convergence of Gradient Descent for Over-parameterized Models using Optimal Transport*, in "Advances in Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2018, https://arxiv.org/abs/1805.09545 , https://hal.archives-ouvertes.fr/hal-01798792

[26] A. DÉFOSSEZ, N. ZEGHIDOUR, N. USUNIER, L. BOTTOU, F. BACH.*SING: Symbol-to-Instrument Neural Generator*, in "Conference on Neural Information Processing Systems (NIPS)", Montréal, Canada, December 2018, https://arxiv.org/abs/1810.09785 , https://hal.archives-ouvertes.fr/hal-01899949

[27] R. M. GOWER, N. LE ROUX, F. BACH.*Tracking the gradients using the Hessian: A new look at variance reducing stochastic methods*, in "International Conference on Artificial Intelligence and Statistics (AISTATS)", Canary Islands, Spain, 2018, https://arxiv.org/abs/1710.07462 - 17 pages, 2 figures, 1 table [*DOI :* 10.07462], https://hal.archives-ouvertes.fr/hal-01652152

[28] M. E. HALABI, F. BACH, V. CEVHER.*Combinatorial Penalties: Which structures are preserved by convex relaxations?*, in "AISTATS 2018 - 22nd International Conference on Artificial Intelligence and Statistics", Canary Islands, Spain, April 2018, https://arxiv.org/abs/1710.06273 [*DOI :* 10.06273], https://hal.archives-ouvertes.fr/hal-01652151

[29] T. KERDREUX, F. PEDREGOSA, A. D'ASPREMONT.*Frank-Wolfe with Subsampling Oracle*, in "ICML 2018 - 35th International Conference on Machine Learning", Stockholm, Sweden, July 2018, https://arxiv.org/abs/1803.07348 , https://hal.archives-ouvertes.fr/hal-01927391

[30] A. KUNDU, F. BACH, C. BHATTACHARYYA.*Convex optimization over intersection of simple sets: improved convergence rate guarantees via an exact penalty approach*, in "AISTATS 2018 - 22nd International Conference on Artificial Intelligence and Statistics", Canary Islands, Spain, April 2018, https://arxiv.org/abs/1710.06465 [*DOI :* 10.06465], https://hal.archives-ouvertes.fr/hal-01652149

[31] E. PAUWELS, F. BACH, J.-P. VERT.*Relating Leverage Scores and Density using Regularized Christoffel Functions*, in "Neural Information Processing Systems", Montréal, Canada, December 2018, https://hal.archives-ouvertes.fr/hal-01796591

[32] L. PILLAUD-VIVIEN, A. RUDI, F. BACH.*Exponential convergence of testing error for stochastic gradient methods*, in "Conference on Learning Theory (COLT)", Stockholm, Sweden, July 2018, https://arxiv.org/abs/1712.04755 , https://hal.archives-ouvertes.fr/hal-01662278

[33] L. PILLAUD-VIVIEN, A. RUDI, F. BACH.*Statistical Optimality of Stochastic Gradient Descent on Hard Learning Problems through Multiple Passes*, in "Neural Information Processing Systems (NeurIPS)", Montréal, Canada, December 2018, https://arxiv.org/abs/1805.10074 , https://hal.archives-ouvertes.fr/hal-01799116

[34] S. J. REDDI, M. ZAHEER, S. SRA, B. POCZOS, F. BACH, R. SALAKHUTDINOV, A. J. SMOLA.*A Generic Approach for Escaping Saddle points*, in "AISTATS 2018 - 22nd International Conference on Artificial Intelligence and Statistics", Canary Islands, Spain, April 2018, https://arxiv.org/abs/1709.01434 , https://hal.archives-ouvertes.fr/hal-01652150

[35] K. SCAMAN, F. BACH, S. BUBECK, Y. T. LEE, L. MASSOULIÉ.*Optimal Algorithms for Non-Smooth Distributed Optimization in Networks*, in "Advances In Neural Information Processing systems", Montreal, Canada, December 2018, https://arxiv.org/abs/1806.00291 - 17 pages, https://hal.archives-ouvertes.fr/hal-01957013

[36] D. SCIEUR, E. OYALLON, A. D'ASPREMONT, F. BACH.*Nonlinear Acceleration of CNNs*, in "ICLR Workshop track", Vancouver, Canada, April 2018, https://hal.archives-ouvertes.fr/hal-01805251

[37] N. TRIPURANENI, N. FLAMMARION, F. BACH, M. I. JORDAN.*Averaging Stochastic Gradient Descent on Riemannian Manifolds*, in "Computational Learning Theory (COLT)", Stockholm, Sweden, July 2018, https://arxiv.org/abs/1802.09128 - COLT 2018, https://hal.archives-ouvertes.fr/hal-01957015

## Other Publications

[38] M. BARRÉ, A. D'ASPREMONT.M*-*Regularized Dictionary Learning*, October 2018, https://arxiv.org/abs/1810.02748 - working paper or preprint [*DOI :* 10.02748], https://hal.archives-ouvertes.fr/hal-01897496

[39] R. BERTHIER, F. BACH, P. GAILLARD.*Gossip of Statistical Observations using Orthogonal Polynomials*, May 2018, https://arxiv.org/abs/1805.08531 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01797016

[40] R. BOLLAPRAGADA, D. SCIEUR, A. D'ASPREMONT.*Nonlinear Acceleration of Momentum and Primal-Dual Algorithms*, October 2018, https://arxiv.org/abs/1810.04539 - working paper or preprint [*DOI :* 10.04539], https://hal.archives-ouvertes.fr/hal-01893921

[41] L. CHIZAT, F. BACH.*A Note on Lazy Training in Supervised Differentiable Programming*, December 2018, https://arxiv.org/abs/1812.07956 - working paper or preprint, https://hal.inria.fr/hal-01945578

[42] C. CILIBERTO, F. BACH, A. RUDI.*Localized Structured Prediction*, December 2018, https://arxiv.org/abs/1806.02402 - 53 pages, 7 figures, 1 algorithm, https://hal.inria.fr/hal-01958863

[43] A. DIEULEVEUT, A. DURMUS, F. BACH.*Bridging the Gap between Constant Step Size Stochastic Gradient Descent and Markov Chains*, April 2018, https://arxiv.org/abs/1707.06386 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01565514

[44] Y. DRORI, A. B. TAYLOR.*Efficient First-order Methods for Convex Minimization: a Constructive Approach*, October 2018, https://arxiv.org/abs/1803.05676 - Code available at https://github.com/AdrienTaylor/GreedyMethods, https://hal.inria.fr/hal-01902048

[45] P. GAILLARD, S. GERCHINOVITZ, M. HUARD, G. STOLTZ.*Uniform regret bounds over $R^d$ for the sequential linear regression problem with the square loss*, February 2018, https://arxiv.org/abs/1805.11386 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01802004

[46] P. GAILLARD, O. WINTENBERGER.*Efficient online algorithms for fast-rate regret bounds under sparsity*, May 2018, https://arxiv.org/abs/1805.09174 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01798201

[47] H. HENDRIKX, F. BACH, L. MASSOULIÉ.*Accelerated decentralized optimization with local updates for smooth and strongly convex objectives*, October 2018, working paper or preprint, https://hal.inria.fr/hal-01893568

[48] T. KERDREUX, A. D'ASPREMONT, S. POKUTTA.*Restarting Frank-Wolfe*, October 2018, https://arxiv.org/abs/1810.02429 - working paper or preprint [*DOI :* 10.02429], https://hal.archives-ouvertes.fr/hal-01893922

[49] A. NOWAK-VILA, F. BACH, A. RUDI.*Sharp Analysis of Learning with Discrete Losses*, October 2018, https://arxiv.org/abs/1810.06839 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01893006

[50] D. M. OSTROVSKII, F. BACH.*Finite-sample Analysis of M-estimators using Self-concordance*, October 2018, https://arxiv.org/abs/1810.06838 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01895127

[51] A. RECANATI, T. KERDREUX, A. D'ASPREMONT.*Reconstructing Latent Orderings by Spectral Clustering*, July 2018, https://arxiv.org/abs/1807.07122 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01846269

[52] A. RECANATI, N. SERVANT, J.-P. VERT, A. D'ASPREMONT.*Robust Seriation and Applications to Cancer Genomics*, July 2018, https://arxiv.org/abs/1806.00664 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01851960

[53] E. K. RYU, A. B. TAYLOR, C. BERGELING, P. GISELSSON.*Operator Splitting Performance Estimation: Tight contraction factors and optimal parameter selection*, December 2018, https://arxiv.org/abs/1812.00146 - working paper or preprint, https://hal.inria.fr/hal-01943622

[54] D. SCIEUR, E. OYALLON, A. D'ASPREMONT, F. BACH.*Nonlinear Acceleration of Deep Neural Networks*, May 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01799269

[55] J. TANG, M. GOLBABAEE, F. BACH, M. E. DAVIES.*Structure-Adaptive Accelerated Coordinate Descent*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01889990

[56] T.-H. VU, A. OSOKIN, I. LAPTEV.*Tube-CNN: Modeling temporal evolution of appearance for object detection in video*, January 2019, https://arxiv.org/abs/1812.02619 - 13 pages, 8 figures, technical report, https://hal.archives-ouvertes.fr/hal-01980339

# Project-Team VALDA

## Value from Data

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Data and Knowledge Representation and Processing**

# Table of contents

# Project-Team VALDA

*Creation of the Team: 2016 December 01, updated into Project-Team: 2018 January 01*

**Keywords:**

**Computer Science and Digital Science:**

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
A3.1.10. - Heterogeneous data
A3.1.11. - Structured data
A3.2. - Knowledge
A3.2.1. - Knowledge bases
A3.2.2. - Knowledge extraction, cleaning
A3.2.3. - Inference
A3.2.4. - Semantic Web
A3.2.5. - Ontologies
A3.2.6. - Linked data
A3.3.2. - Data mining
A3.4.3. - Reinforcement learning
A3.4.5. - Bayesian methods
A3.5.1. - Analysis of large graphs
A4.7. - Access control
A7.2. - Logic in Computer Science
A7.3. - Calculability and computability
A9.1. - Knowledge
A9.8. - Reasoning

**Other Research Topics and Application Domains:**

B6.3.1. - Web
B6.3.4. - Social Networks
B6.5. - Information systems
B9.5.6. - Data science
B9.6.5. - Sociology
B9.6.10. - Digital humanities
B9.7.2. - Open data
B9.9. - Ethics
B9.10. - Privacy

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Serge Abiteboul [Inria, Senior Researcher, until Jan 2018, HDR]
Luc Segoufin [Inria, Senior Researcher, HDR]
Michael Thomazo [Inria, Researcher, from Apr 2018]
Camille Bourgaux [CNRS, Researcher, from Oct 2018]
Olivier Cappe [CNRS, Senior Researcher, from Feb 2018]

**Faculty Member**

Pierre Senellart [Team leader, École normale supérieure, Professor, HDR]

**External Collaborators**

Serge Abiteboul [ARCEP, Board Member, from Feb 2018, HDR]
Yann Ramusat [École normale supérieure, student on a long-term project, until Aug 2018]

**PhD Students**

Julien Grange [École normale supérieure]
Miyoung Han [Télécom ParisTech, until Aug 2018]
Quentin Lobbe [Télécom ParisTech, until Nov 2018]
Mikael Monet [Télécom ParisTech, until Oct 2018]
Karima Rafes [BorderCloud]
Yann Ramusat [École normale supérieure, from Sep 2018]
Alexandre Vigny [Université Paris Diderot, until Sep 2018]
Russac Yoan [École normale supérieure, from Dec 2018]

**Post-Doctoral Fellow**

Nathan Grosshans [École normale supérieure, from Sep 2018, temporary research and teaching assistant (ATER)]

**Visiting Scientist**

Victor Vianu [UCSD & École normale supérieure, from Jun 2018 until Sep 2018]

**Administrative Assistants**

Meriem Henni [Inria, from Apr 2018]
Sandrine Vergès [Inria, until Jun 2018]

# 2. Overall Objectives

## 2.1. Objectives

> Valda's focus is on both *foundational and systems aspects of* complex *data management*, especially *human-centric data*. The data we are interested in is typically heterogeneous, massively distributed, rapidly evolving, intensional, and often subjective, possibly erroneous, imprecise, incomplete. In this setting, Valda is in particular concerned with the optimization of complex resources such as computer time and space, communication, monetary, and privacy budgets. The goal is to extract *value from data*, beyond simple query answering.

Data management [43], [55] is now an old, well-established field, for which many scientific results and techniques have been accumulated since the sixties. Originally, most works dealt with static, homogeneous, and precise data. Later, works were devoted to heterogeneous data [40] [45], and possibly distributed [91] but at a small scale.

However, these classical techniques are poorly adapted to handle the new challenges of data management. Consider human-centric data, which is either produced by humans, e.g., emails, chats, recommendations, or produced by systems when dealing with humans, e.g., geolocation, business transactions, results of data analysis. When dealing with such data, and to accomplish any task to extract value from such data, we rapidly encounter the following facets:

- *Heterogeneity*: data may come in many different structures such as unstructured text, graphs, data streams, complex aggregates, etc., using many different schemas or ontologies.
- *Massive distribution*: data may come from a large number of autonomous sources distributed over the web, with complex access patterns.
- *Rapid evolution*: many sources may be producing data in real time, even if little of it is perhaps relevant to the specific application. Typically, recent data is of particular interest and changes have to be monitored.
- *Intensionality*[0]: in a classical database, all the data is available. In modern applications, the data is more and more available only intensionally, possibly at some cost, with the difficulty to discover which source can contribute towards a particular goal, and this with some uncertainty.
- *Confidentiality and security*: some personal data is critical and need to remain confidential. Applications manipulating personal data must take this into account and must be secure against linking.
- *Uncertainty*: modern data, and in particular human-centric data, typically includes errors, contradictions, imprecision, incompleteness, which complicates reasoning. Furthermore, the subjective nature of the data, with opinions, sentiments, or biases, also makes reasoning harder since one has, for instance, to consider different agents with distinct, possibly contradicting knowledge.

These problems have already been studied individually and have led to techniques such as *query rewriting* [71] or *distributed query optimization* [77].

Among all these aspects, intensionality is perhaps the one that has least been studied, so we will pay particular attention to it. Consider a user's query, taken in a very broad sense: it may be a classical database query, some information retrieval search, a clustering or classification task, or some more advanced knowledge extraction request. Because of intensionality of data, solving such a query is a typically dynamic task: each time new data is obtained, the partial knowledge a system has of the world is revised, and query plans need to be updated, as in adaptive query processing [61] or aggregated search [90]. The system then needs to decide, based on this partial knowledge, of the best next access to perform. This is reminiscent of the central problem of reinforcement learning [88] (train an agent to accomplish a task in a partially known world based on rewards obtained) and of active learning [85] (decide which action to perform next in order to optimize a learning strategy) and we intend to explore this connection further.

Uncertainty of the data interacts with its intensionality: efforts are required to obtain more precise, more complete, sounder results, which yields a trade-off between *processing cost* and *data quality*.

Other aspects, such as heterogeneity and massive distribution, are of major importance as well. A standard data management task, such as query answering, information retrieval, or clustering, may become much more challenging when taking into account the fact that data is not available in a central location, or in a common format. We aim to take these aspects into account, to be able to apply our research to real-world applications.

## 2.2. The Issues

We intend to tackle hard technical issues such as query answering, data integration, data monitoring, verification of data-centric systems, truth finding, knowledge extraction, data analytics, that take a different flavor in this modern context. In particular, we are interested in designing strategies to *minimize data access cost towards a specific goal, possibly a massive data analysis task*. That cost may be in terms of communication (accessing data in distributed systems, on the Web), of computational resources (when data is produced

[0]We use the spelling *intensional*, as in mathematical logic and philosophy, to describe something that is neither available nor defined in *extension*; *intensional* is derived from *intension*, while *intentional* is derived from *intent*.

by complex tools such as information extraction, machine learning systems, or complex query processing), of monetary budget (paid-for application programming interfaces, crowdsourcing platforms), or of a privacy budget (as in the standard framework of differential privacy).

A number of data management tasks in Valda are inherently intractable. In addition to properly characterizing this intractability in terms of complexity theory, we intend to develop solutions for solving these tasks in practice, based on approximation strategies, randomized algorithms, enumeration algorithms with constant delay, or identification of restricted forms of data instances lowering the complexity of the task.

# 3. Research Program

## 3.1. Scientific Foundations

We now detail some of the scientific foundations of our research on complex data management. This is the occasion to review connections between data management, especially on complex data as is the focus of Valda, with related research areas.

### *3.1.1. Complexity & Logic.*

Data management has been connected to logic since the advent of the relational model as main representation system for real-world data, and of first-order logic as the logical core of database querying languages [43]. Since these early developments, logic has also been successfully used to capture a large variety of query modes, such as data aggregation [76], recursive queries (Datalog), or querying of XML databases [55]. Logical formalisms facilitate reasoning about the expressiveness of a query language or about its complexity.

The main problem of interest in data management is that of query evaluation, i.e., computing the results of a query over a database. The complexity of this problem has far-reaching consequences. For example, it is because first-order logic is in the $AC_0$ complexity class that evaluation of SQL queries can be parallelized efficiently. It is usual [89] in data management to distinguish *data complexity*, where the query is considered to be fixed, from *combined complexity*, where both the query and the data are considered to be part of the input. Thus, though conjunctive queries, corresponding to a simple SELECT-FROM-WHERE fragment of SQL, have PTIME data complexity, they are NP-hard in combined complexity. Making this distinction is important, because data is often far larger (up to the order of terabytes) than queries (rarely more than a few hundred bytes). Beyond simple query evaluation, a central question in data management remains that of complexity; tools from algorithm analysis, and complexity theory can be used to pinpoint the tractability frontier of data management tasks.

### *3.1.2. Automata Theory.*

Automata theory and formal languages arise as important components of the study of many data management tasks: in temporal databases [42], queries, expressed in temporal logics, can often by compiled to automata; in graph databases [51], queries are naturally given as automata; typical query and schema languages for XML databases such as XPath and XML Schema can be compiled to tree automata [81], or for more complex languages to data tree automata[4]. Another reason of the importance of automata theory, and tree automata in particular, comes from Courcelle's results [59] that show that very expressive queries (from the language of monadic second-order language) can be evaluated as tree automata over *tree decompositions* of the original databases, yielding linear-time algorithms (in data complexity) for a wide variety of applications.

### *3.1.3. Verification.*

Complex data management also has connections to verification and static analysis. Besides query evaluation, a central problem in data management is that of deciding whether two queries are *equivalent* [43]. This is critical for query optimization, in order to determine if the rewriting of a query, maybe cheaper to evaluate, will return the same result as the original query. Equivalence can easily be seen to be an instance of the problem of (non-)satisfiability: $q \equiv q'$ if and only if $(q \wedge \neg q') \vee (\neg q \wedge q')$ is not satisfiable. In other words, some aspects of query optimization are static analysis issues. Verification is also a critical part of any database application where it is important to ensure that some property will never (or always) arise [57].

### *3.1.4. Workflows.*

The orchestration of distributed activities (under the responsibility of a conductor) and their choreography (when they are fully autonomous) are complex issues that are essential for a wide range of data management applications including notably, e-commerce systems, business processes, health-care and scientific workflows. The difficulty is to guarantee consistency or more generally, quality of service, and to statically verify critical properties of the system. Different approaches to workflow specifications exist: automata-based, logic-based, or predicate-based control of function calls [39].

### *3.1.5. Probability & Provenance.*

To deal with the uncertainty attached to data, proper models need to be used (such as attaching *provenance* information to data items and viewing the whole database as being *probabilistic*) and practical methods and systems need to be developed to both reliably estimate the uncertainty in data items and properly manage provenance and uncertainty information throughout a long, complex system.

The simplest model of data uncertainty is the NULLs of SQL databases, also called Codd tables [43]. This representation system is too basic for any complex task, and has the major inconvenient of not being closed under even simple queries or updates. A solution to this has been proposed in the form of *conditional tables* [73] where every tuple is annotated with a Boolean formula over independent Boolean random events. This model has been recognized as foundational and extended in two different directions: to more expressive models of *provenance* than what Boolean functions capture, through a semiring formalism [69], and to a probabilistic formalism by assigning independent probabilities to the Boolean events [70]. These two extensions form the basis of modern provenance and probability management, subsuming in a large way previous works [58], [52]. Research in the past ten years has focused on a better understanding of the tractability of query answering with provenance and probabilistic annotations, in a variety of specializations of this framework [87] [75], [48].

### *3.1.6. Machine Learning.*

Statistical machine learning, and its applications to data mining and data analytics, is a major foundation of data management research. A large variety of research areas in complex data management, such as wrapper induction [83], crowdsourcing [50], focused crawling [68], or automatic database tuning [53] critically rely on machine learning techniques, such as classification [72], probabilistic models [67], or reinforcement learning [88].

Machine learning is also a rich source of complex data management problems: thus, the probabilities produced by a conditional random field [78] system result in probabilistic annotations that need to be properly modeled, stored, and queried.

Finally, complex data management also brings new twists to some classical machine learning problems. Consider for instance the area of *active learning* [85], a subfield of machine learning concerned with how to optimally use a (costly) oracle, in an interactive manner, to label training data that will be used to build a learning model, e.g., a classifier. In most of the active learning literature, the cost model is very basic (uniform or fixed-value costs), though some works [84] consider more realistic costs. Also, oracles are usually assumed to be perfect with only a few exceptions [62]. These assumptions usually break when applied to complex data management problems on real-world data, such as crowdsourcing.

Having situated Valda's research area within its broader scientific scope, we now move to the discussion of Valda's application domains.

## 3.2. Research Directions

We now detail three main research axes within the research agenda of Valda. For each axis, we first mention the leading researcher, and other permanent members involved.

### 3.2.1. *Foundations of data management (Luc Segoufin; Serge Abiteboul, Camille Bourgaux, Michaël Thomazo, Pierre Senellart).*

Foundations of data management

The systems we are interested in, i.e., for manipulating heterogeneous and confidential data, rapidly changing and massively distributed, are inherently error-prone. The need for formal methods to verify data management systems is best illustrated by the long list of famous leakages of sensitive or personal data that made the front pages of newspapers recently. Moreover, because of the cost in accessing intensional data, it is important to optimize the resources needed for manipulating them.

This creates a need for solid and high-level foundations of DBMS in a manner that is easier to understand, while also facilitating optimization and verification of its critical properties.

In particular these foundations are necessary for various design and reasoning tasks. It allows for clean specifications of key properties of the system such as confidentiality, access control, robustness etc. Once clean specifications are available, it opens the door for formal and runtime verification of the specification. It also permits the design of appropriate query languages – with good expressive power, with limited usage of resources –, the design of good indexes – for optimized evaluation –, and so on. Note that access control policies currently used in database management systems are relatively crude – for example, PostgreSQL offers access control rules on tables, views, or tuples (*row security policies*), but provides no guarantee that these access methods do not contradict each other, or that a user may have access through a query to information that she is not supposed to have access to.

Valda involves leading researchers in the formal verification of data flow in a system manipulating data. Other notable teams involve the WAVE project [0] at U. C. San Diego, and the Business Artifact [0] research program of IBM. One of Valda's objectives is to continue this line of research.

In the short run, we plan to contribute to the state of the art of foundations of systems manipulating data by identifying new scenarios, i.e., specification formalisms, query languages, index structures, query evaluation plans, etc., that allow for any of the tasks mentioned above: formal or runtime verification, optimization etc. Several such scenarios are already known and Valda researchers contributed significantly to their discovery [57], [74], [64], but this research is still in infancy and there is a clear need for more functionalities and more efficiency. This research direction has many facets.

One of the facet is to develop new logical frameworks and new automaton models, with good algorithmic properties (for instance efficient emptiness test, efficient inclusion test and so on), in order to develop a toolbox for reasoning task around systems manipulating data. This toolbox can then be used for higher level tasks such as optimization, verification [57], or query rewriting using views [64].

Another facet is to develop new index structures and new algorithms for efficient query evaluation. For example the enumeration of the output of a query requires the construction of index structures allowing for efficient compressed representation of the output with efficient streaming decompression algorithms as we aim for a constant delay between any two consecutive outputs [82]. We have contributed a lot to this fields by providing several such indexes [74] but there remains a lot to be investigated.

Our medium-term goal is to investigate the borders of feasibility of all the reasoning tasks above. For instance what are the assumptions on data that allow for computable verification problems? When is it not possible at all? When can we hope for efficient query answering, when is it hopeless? This is a problem of theoretical nature which is necessary for understanding the limit of the methods and driving research towards the scenarios where positive results may be obtainable.

A typical result would be to show that constant delay enumeration of queries is not possible unless the database verify property A and the query property B. Another typical result would be to show that having a robust access control policy verifying at the same time this and that property is not achievable.

[0] http://db.ucsd.edu/WAVE/default.html
[0] http://researcher.watson.ibm.com/researcher/view_group.php?id=2501

Very few such results exist nowadays. If many problems are shown undecidable or decidable, charting the frontier of tractability (say linear time) remains a challenge.

Only when we will have understood the limitation of the method (medium-term goal) and have many examples where this is possible, we can hope to design a solid foundation that allowing for a good trade-off between what can be done (needs from the users) and what can be achieved (limitation from the system). This will be our long-term goal.

### *3.2.2. Uncertainty and provenance of data (Pierre Senellart; Camille Bourgaux, Olivier Cappé, Michaël Thomazo, Luc Segoufin).*

Uncertainty and provenance of data

This research axis deals with the modeling and efficient management of data that come with some uncertainty (probabilistic distributions, logical incompleteness, missing values, open-world assumption, etc.) and with provenance information (indicating where the data originates from), as well as with the extraction of uncertainty and provenance annotations from real-world data. Interestingly, the foundations and tools for uncertainty management often rely on provenance annotations. For example, a typical way to compute the probability of query results in probabilistic databases is first to generate the provenance of these query results (in some Boolean framework, e.g., that of Boolean functions or of provenance semirings), and then to compute the probability of the resulting provenance annotation. For this reason, we will deal with uncertainty and provenance in a unified manner.

Valda researchers have carried out seminal work on probabilistic databases [75], [44][7], provenance management [47], incomplete information [46], and uncertainty analysis and propagation in conflicting datasets [65], [41]. These research areas have reached a point where the foundations are well-understood, and where it becomes critical, while continuing developing the theory of uncertain and provenance data management, to move to concrete implementations and applications to real-world use cases.

In the short term, we will focus on implementing techniques from the database theory literature on provenance and uncertainty data management, in the direction of building a full-featured database management add-on that transparently manages provenance and probability annotations for a large class of querying tasks. This work has started recently with the creation of the ProvSQL extension to PostgreSQL, discussed in more details in the following section. To support this development work, we need to resolve the following research question: what representation systems and algorithms to use to support both semiring provenance frameworks [69], extensions to queries with negation [66], aggregation [49], or recursion [80]?

Next, we will study how to add support for incompleteness, probabilities, and provenance annotations in the scenarios identified in the first axis, and how to extract and derive such annotations from real-world datasets and tasks. We will also work on the efficiency of our uncertain data management system, and compare it to other uncertainty management solutions, in the perspective of making it a fully usable system, with little overhead compared to a classical database management system. This requires a careful choice of the provenance representation system used, which should be both compact and amenable to probability computations. We will study practical applications of uncertainty management. As an example, we intend to consider routing in public transport networks, given a probabilistic model on the reliability and schedule uncertainty of different transit routes. The system should be able to provide a user with itinerary to get to have a (probabilistic) guarantee to be at its destination within a given time frame, which may not be the shortest route in the classical sense.

One overall long-term goal is to reach a full understanding of the interactions between query evaluation or other broader data management tasks and uncertain and annotated data models. We would in particular want to go towards a full classification of tractable (typically polynomial-time) and intractable (typically NP-hard for decision problems, or #P-hard for probability evaluation) tasks, extending and connecting the query-based dichotomy [60] on probabilistic query evaluation with the instance-based one of [47], [48].

Another long-term goal is to consider more dynamic scenarios than what has been considered so far in the uncertain data management literature: when following a workflow, or when interacting with intensional data

sources, how to properly represent and update uncertainty annotations that are associated with data. This is critical for many complex data management scenarios where one has to maintain a probabilistic current knowledge of the world, while obtaining new knowledge by posing queries and accessing data sources. Such intensional tasks requires minimizing jointly data uncertainty and cost to data access.

### *3.2.3. Personal information management (Serge Abiteboul; Pierre Senellart).*

Personal information management

This is a more applied direction of research that will be the context to study issues of interest (see discussion in application domains further).

A typical person today usually has data on several devices and in a number of commercial systems that function as data traps where it is easy to check in information and difficult to remove it or sometimes to simply access it. It is also difficult, sometimes impossible, to control data access by other parties. This situation is unsatisfactory because it requires users to trade privacy against convenience but also, because it limits the value we, as individuals and as a society, can derive from the data. This leads to the concept of Personal Information Management System, in short, a Pims.

A Pims runs, on a user's server, the services selected by the user, storing and processing the user's data. The Pims centralizes the user's personal information. It is a digital home. The Pims is also able to exert control over information that resides in external services (for example, Facebook), and that only gets replicated inside the Pims. See, for instance, [38] for a discussion on the advantages of Pims, as well as issues they raise, e.g. security issues. It is argued there that the main reason for a user to move to Pims is these systems enable great new functionalities.

Valda will study in particular the integration of the user's data. Researchers in the team have already provided important contributions in the context of data integration, notably in the context of the Webdam ERC (2009–2013).

Based on such an integration, Pims can provide a functions, that goes beyond simple query answering:

- Global search over the person's data with a semantic layer using a personal ontology (for example, the data organization the person likes and the person's terminology for data) that helps give meaning to the data;
- Automatic synchronization of data on different devices/systems, and global task sequencing to facilitate interoperating different devices/services;
- Exchange of information and knowledge between "friends" in a truly social way, even if these use different social network platforms, or no platform at all;
- Centralized control point for connected objects, a hub for the Internet of Things; and
- Data analysis/mining over the person's information.

The focus on personal data and these various aspects raise interesting technical challenges that we intend to address.

In the short term, we intend to continue work on the ThymeFlow system to turn it into an easily extendable and deployable platform for the management of personal information – we will in particular encourage students from the M2 *Web Data Management* class taught by Serge and Pierre in the MPRI programme to use this platform in their course projects. The goal is to make it easy to add new functionalities (such as new source *synchronizers* to retrieve data and propagate updates to original data sources, and *enrichers* to add value to existing data) to considerably broaden the scope of the platform and consequently expand its value.

In the medium term, we will continue the work already started that focuses in turning information into knowledge and in knowledge integration. Issues related to intensionality or uncertainty will in particular be considered, relying on the works produced in the other two research axes. We stress, in particular, the importance of minimizing the cost to data access (or, in specific scenarios, the privacy cost associated with obtaining data items) in the context of personal information management: legacy data is often only available through costly APIs, interaction between several Pims may require sharing information within a strict privacy budget, etc. For these reasons, intensionality of data will be a strong focus of the research.

In the long term, we intend to use the knowledge acquired and machine learning techniques to predict the user's behavior and desires, and support new digital assistant functions, providing real *value from data*. We will also look into possibilities for deploying the ThymeFlow platform at a large scale, perhaps in collaboration with industry partners.

# 4. Application Domains

## 4.1. Personal Information Management Systems

We recall that Valda's focus is on human-centric data, i.e., data produced by humans, explicitly or implicitly, or more generally containing information about humans. Quite naturally, we will use as a privileged application area to validate Valda's results that of personal information management systems (Pims for short) [38].

A Pims is a system that allows a user to integrate her own data, e.g., emails and other kinds of messages, calendar, contacts, web search, social network, travel information, work projects, etc. Such information is commonly spread across different services. The goal is to give back to a user the control on her information, allowing her to formulate queries such as "What kind of interaction did I have recently with Alice B.?", "Where were my last ten business trips, and who helped me plan them?". The system has to orchestrate queries to the various services (which means knowing the existence of these services, and how to interact with them), integrate information from them (which means having data models for this information and its representation in the services), e.g., align a GPS location of the user to a business address or place mentioned in an email, or an event in a calendar to some event in a Web search. This information must be accessed intensionally: for instance, costly information extraction tools should only be run on emails which seem relevant, perhaps identified by a less costly cursory analysis (this means, in turn, obtaining a cost model for access to the different services). Impacted people can be found by examining events in the user's calendar and determining who is likely to attend them, perhaps based on email exchanges or former events' participant lists. Of course, uncertainty has to be maintained along the entire process, and provenance information is needed to explain query results to the user (e.g., indicate which meetings and trips are relevant to each person of the output). Knowledge about services, their data models, their costs, need either to be provided by the system designer, or to be automatically learned from interaction with these services, as in [83].

One motivation for that choice is that Pims concentrate many of the problems we intend to investigate: heterogeneity (various sources, each with a different structure), massive distribution (information spread out over the Web, in numerous sources), rapid evolution (new data regularly added), intensionality (knowledge from Wikidata, OpenStreetMap...), confidentiality and security (mostly private data), and uncertainty (very variable quality). Though the data is distributed, its size is relatively modest; other applications may be considered for works focusing on processing data at large scale, which is a potential research direction within Valda, though not our main focus. Another strong motivation for the choice of Pims as application domain is the importance of this application from a societal viewpoint.

A Pims is essentially a system built on top of a user's *personal knowledge base*; such knowledge bases are reminiscent of those found in the Semantic Web, e.g., linked open data. Some issues, such as ontology alignment [86] exist in both scenarios. However, there are some fundamental differences in building personal knowledge bases vs collecting information from the Semantic Web: first, the scope is quite smaller, as one is only interested in knowledge related to a given individual; second, a small proportion of the data is already present in the form of semantic information, most needs to be extracted and annotated through appropriate wrappers and enrichers; third, though the linked open data is meant to be read-only, the only update possible to a user being adding new triples, a personal knowledge base is very much something that a user needs to be able to edit, and propagating updates from the knowledge base to original data sources is a challenge in itself.

## 4.2. Web Data

The choice of Pims is not exclusive. We intend to consider other application areas as well. In particular, we have worked in the past and have a strong expertise on Web data [45] in a broad sense: semi-structured, structured, or unstructured content extracted from Web databases [83]; knowledge bases from the Semantic Web [86]; social networks [79]; Web archives and Web crawls [63]; Web applications and deep Web databases [56]; crowdsourcing platforms [50]. We intend to continue using Web data as a natural application domain for the research within Valda when relevant. For instance [54], deep Web databases are a natural application scenario for intensional data management issues: determining if a deep Web database contains some information requires optimizing the number of costly requests to that database.

A common aspect of both personal information and Web data is that their exploitation raises ethical considerations. Thus, a user needs to remain fully in control of the usage that is made of her personal information; a search engine or recommender system that ranks Web content for display to a specific user needs to do so in an unbiased, justifiable, manner. These ethical constraints sometimes forbid some technically solutions that may be technically useful, such as sharing a model learned from the personal data of a user to another user, or using blackboxes to rank query result. We fully intend to consider these ethical considerations within Valda. One of the main goals of a Pims is indeed to empower the user with a full control on the use of this data.

# 5. New Software and Platforms

## 5.1. ProvSQL

KEYWORDS: Databases - Provenance - Probability
FUNCTIONAL DESCRIPTION: The goal of the ProvSQL project is to add support for (m-)semiring provenance and uncertainty management to PostgreSQL databases, in the form of a PostgreSQL extension/module/plugin.
NEWS OF THE YEAR: Support for where-provenance has been completed. Numerous bug fixes. A docker version has been produced, for ease of installation. Demonstration scenarios are included.

- Participants: Pierre Senellart and Yann Ramusat
- Contact: Pierre Senellart
- Publications: Provenance and Probabilities in Relational Databases: From Theory to Practice - ProvSQL: Provenance and Probability Management in PostgreSQL
- URL: https://github.com/PierreSenellart/provsql

## 5.2. WAE

*Web archive explorer*
KEYWORDS: Information extraction - Web archives
FUNCTIONAL DESCRIPTION: The Web archive explorer is a system for extracting, fragmenting and exploring Web archives.

- Contact: Quentin Lobbe
- Publications: Archives, Web fragments and diasporas. For a disaggregated exploration of web archives related to online representations of diasporas - Where the dead blogs are. A Disaggregated Exploration of Web Archives to Reveal Extinct Online Collectives - Revealing Historical Events out of Web Archives
- URL: https://github.com/lobbeque/archive-miner

## 5.3. apxproof

KEYWORD: LaTeX

FUNCTIONAL DESCRIPTION: apxproof is a LaTeX package facilitating the typesetting of research articles with proofs in appendix, a common practice in database theory and theoretical computer science in general. The appendix material is written in the LaTeX code along with the main text which it naturally complements, and it is automatically deferred. The package can automatically send proofs to the appendix, can repeat in the appendix the theorem environments stated in the main text, can section the appendix automatically based on the sectioning of the main text, and supports a separate bibliography for the appendix material.

RELEASE FUNCTIONAL DESCRIPTION: Numerous bug fixes and robustness enhancements, link theorems to their repeated Versions, proper management of equations in repeated theorems
NEWS OF THE YEAR: Major 1.1.0 release adding several features (link theorems to their repeated versions, proper management of equations in repeated theorems), beyond this, bug fixes, robustness enhancements, better support for some document classes.

- Participant: Pierre Senellart
- Contact: Pierre Senellart
- URL: https://github.com/PierreSenellart/apxproof

## 5.4. Sgvizler2

KEYWORDS: SPARQL - Data visualization - JavaScript
FUNCTIONAL DESCRIPTION: This project is the reboot in Typescript of project Sgvizler of Martin G. Skjæveland.

- Partners: LRI - Laboratoire de Recherche en Informatique - BorderCloud
- Contact: Karima Rafes
- URL: https://github.com/BorderCloud/sgvizler2

## 5.5. SPARQL-PHP

KEYWORDS: SPARQL - PHP
FUNCTIONAL DESCRIPTION: Very simple SPARQL client for PHP.

- Partners: LRI - Laboratoire de Recherche en Informatique - BorderCloud
- Contact: Karima Rafes
- URL: https://github.com/BorderCloud/SPARQL

## 5.6. TFT

*Tester for Triplestore*
KEYWORDS: PHP - SPARQL
FUNCTIONAL DESCRIPTION: TFT (Tester for Triplestore) is a script PHP to pass tests through a SPARQL service.

- Partners: LRI - Laboratoire de Recherche en Informatique - BorderCloud
- Contact: Karima Rafes
- URL: https://github.com/BorderCloud/TFT

# 6. New Results

## 6.1. Query Enumeration

Query enumeration is the problem of enumerating the results of a query over a database one by one; the goal is to obtain, after some initial low preprocessing time (e.g., linear in the data), one solution after the other with low delay (e.g., constant-time) in between.

In a first work [26], we consider the enumeration of MSO queries over strings under updates. For each MSO query we build an index structure enjoying the following properties: The index structure can be constructed in linear time, it can be updated in logarithmic time and it allows for constant delay time enumeration. This improves from the previous known index structures allowing for constant delay enumeration that would need to be reconstructed from scratch, hence in linear time, in the presence of updates. We allow relabeling updates, insertion of individual labels and removal of individual labels.

In a second work [29], we consider the evaluation of first-order queries over classes of databases that are nowhere dense. The notion of nowhere dense classes was introduced by Nešetřil and Ossona de Mendez as a formalization of classes of "sparse" graphs and generalizes many well-known classes of graphs, such as classes of bounded degree, bounded treewidth, or bounded expansion. It has recently been shown by Grohe, Kreutzer, and Siebertz that over nowhere dense classes of databases, first-order sentences can be evaluated in pseudo-linear time (pseudo-linear time means that for all $\varepsilon$ there exists an algorithm working in time $O(n^{1+\varepsilon})$, where n is the size of the database). For first-order queries of higher arities, we show that over any nowhere dense class of databases, the set of their solutions can be enumerated with constant delay after a pseudo-linear time preprocessing. In the same context, we also show that after a pseudo-linear time preprocessing we can, on input of a tuple, test in constant time whether it is a solution to the query.

## 6.2. Provenance Circuits

We are interested in obtaining efficiently compact representation of the provenance of a query over a database.

In [28], we generalize three existing graph algorithms to compute the provenance of regular path queries over graph databases, in the framework of provenance semirings – algebraic structures that can capture different forms of provenance. Each algorithm yields a different trade-off between time complexity and generality, as each requires different properties over the semiring. Together, these algorithms cover a large class of semirings used for provenance (top-k, security, etc.). Experimental results suggest these approaches are complementary and practical for various kinds of provenance indications, even on a relatively large transport network.

In [16], we showcase ProvSQL, an open-source module for the PostgreSQL database management system that adds support for computation of provenance and probabilities of query results. A large range of provenance formalisms are supported, including all those captured by provenance semirings, provenance semirings with monus, as well as where-provenance. Probabilistic query evaluation is made possible through the use of knowledge compilation tools, in addition to standard approaches such as enumeration of possible worlds and Monte-Carlo sampling. ProvSQL supports a large subset of non-aggregate SQL queries.

Finally, in [20], [35], we focus on knowledge compilation, which can be used to obtain compact circuit-based representations of (Boolean) provenance. Some width parameters of the circuit, such as bounded treewidth or pathwidth, can be leveraged to convert the circuit to structured classes, e.g., deterministic structured NNFs (d-SDNNFs) or OBDDs. We show how to connect the width of circuits to the size of their structured representation, through upper and lower bounds. For the upper bound, we show how bounded-treewidth circuits can be converted to a d-SDNNF, in time linear in the circuit size. Our bound, unlike existing results, is constructive and only singly exponential in the treewidth. We show a related lower bound on monotone DNF or CNF formulas, assuming a constant bound on the arity (size of clauses) and degree (number of occurrences of each variable). Specifically, any d-SDNNF (resp., SDNNF) for such a DNF (resp., CNF) must be of exponential size in its treewidth; and the same holds for pathwidth when compiling to OBDDs. Our lower bounds, in contrast with most previous work, apply to any formula of this class, not just a well-chosen family. Hence, for our language of DNF and CNF, pathwidth and treewidth respectively characterize the efficiency of compiling to OBDDs and (d-)SDNNFs, that is, compilation is singly exponential in the width parameter.

## 6.3. Exploiting Content from the Web

One of our main domain of application is that of Web content. We investigate methods to acquire and exploit content from the Web.

In [30], we analyze form-based websites to discover sequences of actions and values that result in a valid form submission. Rather than looking at the text or DOM structure of the form, our method is driven by solving constraints involving the underlying client-side JavaScript code. In order to deal with the complexity of client-side code, we adapt a method from program analysis and testing, concolic testing, which mixes concrete code execution, symbolic code tracing, and constraint solving to find values that lead to new code paths. While concolic testing is commonly used for detecting bugs in stand-alone code with developer support, we show how it can be applied to the very different problem of filling Web forms. We evaluate our system on a benchmark of both real and synthetic Web forms.

In [21], we investigate *focused crawling*: collecting as many Web pages relevant to a target topic as possible while avoiding irrelevant pages, reflecting limited resources available to a Web crawler. We improve on the efficiency of focused crawling by proposing an approach based on reinforcement learning. Our algorithm evaluates hyperlinks most profitable to follow over the long run, and selects the most promising link based on this estimation. To properly model the crawling environment as a Markov decision process, we propose new representations of states and actions considering both content information and the link structure. The size of the state-action space is reduced by a generalization process. Based on this generalization, we use a linear-function approximation to update value functions. We investigate the trade-off between synchronous and asynchronous methods. In experiments, we compare the performance of a crawling task with and without learning; crawlers based on reinforcement learning show better performance for various target topics.

Finally, in [23], [24] we propose a framework to follow the dynamics of vanished Web communities, based on the exploration of corpora of Web archives. To achieve this goal, we define a new unit of analysis called Web fragment: a semantic and syntactic subset of a given Web page, designed to increase historical accuracy. This contribution has practical value for those who conduct large-scale archive exploration (in terms of time range and volume) or are interested in computational approaches to Web history and social science.

## 6.4. Knowledge Bases

Knowledge bases are collection of semantic facts (typically of the form subject–predicate–object) along with possible logical rules (e.g., in the form of existential rules) that apply to these facts. We investigate querying, data integration, and inference in such knowledge bases.

In [27], we focus on autocompletion of SPARQL queries over knowledge bases. We analyze several autocompletion features proposed by the main editors, highlighting the needs currently not taken into account while met by a user community we work with, scientists. Second, we introduce the first (to our knowledge) autocompletion approach able to consider snippets (fragments of SPARQL query) based on queries expressed by previous users, enriching the user experience. Third, we introduce a usable, open and concrete solution able to consider a large panel of SPARQL autocompletion features that we have implemented in an editor. Last but not least, we demonstrate the interest of our approach on real biomedical queries involving services offered by the Wikidata collaborative knowledge base.

In [25], we introduce a novel open-source framework for integrating the data of a user from different sources into a single knowledge base. Our framework integrates data of different kinds into a coherent whole, starting with email messages, calendar, contacts, and location history. We show how event periods in the user's location data can be detected and how they can be aligned with events from the calendar. This allows users to query their personal information within and across different dimensions, and to perform analytics over their emails, events, and locations. Our system models data using RDF, extending the schema.org vocabulary and providing a SPARQL interface.

Finally, in [22], [32], we view knowledge bases as composed of an instance that contains incomplete data and a set of existential rules, and investigate ontology-based query answering: answers to queries are logically entailed from the knowledge base. This brings to light the fundamental chase tool, and its different variants that have been proposed in the literature. It is well-known that the problem of determining, given a chase variant and a set of existential rules, whether the chase will halt on a given instance / on any instance, is undecidable. Hence, a crucial issue is whether it becomes decidable for known subclasses of existential rules. We consider

linear existential rules, a simple yet important subclass of existential rules. We study the decidability of the associated chase termination problem for different chase variants, with a novel approach based on a single graph and a single notion of forbidden pattern. Besides the theoretical interest of a unified approach, an original result is the decidability of the restricted chase termination for linear existential rules.

## 6.5. Transparency and Bias

In this last set of results, we investigate transparency and bias in data management.

Bias in online information has recently become a pressing issue, with search engines, social networks and recommendation services being accused of exhibiting some form of bias. In [15], we make the case for a systematic approach towards measuring bias. To this end, we discuss formal measures for quantifying the various types of bias, we outline the system components necessary for realizing them, and we highlight the related research challenges and open problems.

In [19], we pursue an investigation of data-driven collaborative work-flows. In the model, peers can access and update local data, causing side-effects on other peers' data. In this paper, we study means of explaining to a peer her local view of a global run, both at runtime and statically. We consider the notion of "scenario for a given peer" that is a subrun observationally equivalent to the original run for that peer. Because such a scenario can sometimes differ significantly from what happens in the actual run, thus providing a misleading explanation , we introduce and study a faithfulness requirement that ensures closer adherence to the global run. We show that there is a unique minimal faithful scenario, that explains what is happening in the global run by extracting only the portion relevant to the peer. With regard to static explanations, we consider the problem of synthesizing, for each peer, a "view program" whose runs generate exactly the peer's observations of the global runs. Assuming some conditions desirable in their own right, namely transparency and boundedness, we show that such a view program exists and can be synthesized. As an added benefit, the view program rules provide provenance information for the updates observed by the peer.

Finally, in two articles oriented towards applications and policy, we discuss bias and neutrality and their impact on regulation. In [18] we discuss the different forms of neutrality in the digital world, from the neutrality of networks to neutrality of content. In [17], we investigate the impact of bias and neutrality concerns on algorithms used by businesses.

# 7. Partnerships and Cooperations

## 7.1. Regional Initiatives

Michaël Thomazo has obtained a 6k€ budget from the Île-de-France region (DIM RFSI – *Réseau Francilien en Sciences Informatiques*) entitled *ISORE: Indexation sémantique d'ontologies, le cas des règles existentielles.* The grant was awarded when Michaël Thomazo was part of the Inria Saclay Cedar team, but the budget was transferred to the Valda team.

## 7.2. National Initiatives

### *7.2.1. ANR*

Valda has been part of two ANR projects in 2018:

- HEADWORK (budget managed by Inria), together with IRISA (Druid, coordinator), Inria Lille (Links & Spirals), and Inria Rennes (Sumo), and two application partners: MNHN (Cesco) and FouleFactory. The topic is workflows for crowdsourcing. See http://headwork.gforge.inria.fr/.
- BioQOP (budget managed by ENS), with Idemia (coordinator) and GREYC, on the optimization of queries for privacy-aware biometric data management. See http://bioqop.di.ens.fr/.

In addition, two ANR projects were accepted in 2018 and will start early 2019:

- CQFD (budget managed by Inria), with Inria Sophia (GraphIK, coordinator), LaBRI, LIG, Inria Saclay (Cedar), IRISA, Inria Lille (Spirals), and Télécom ParisTech, on complex ontological queries over federated and heterogeneous data.
- QUID (budget managed by Inria), LIGM (coordinator), IRIF, and LaBRI, on incomplete and inconsistent data.

## 7.3. International Initiatives

### *7.3.1. IIL projects*

Valda has strong collaborations with the following international groups:

Univ. Edinburgh, United Kingdom: Peter Buneman and Leonid Libkin

Univ. Oxford, United Kingdom: Michael Benedikt, Evgeny Kharlamov, Dan Olteanu, and Georg Gottlob

TU Dresden, Germany: Markus Krötzsch and Sebastian Rudolph

Dortmund University, Germany: Thomas Schwentick

Warsaw University, Poland: Mikołaj Bojańczyk and Szymon Toruńczyk

Tel Aviv University, Israel: Daniel Deutch and Tova Milo

Drexel University, USA: Julia Stoyanovich

Univ. California San Diego, USA: Victor Vianu

National University of Singapore: Stéphane Bressan

## 7.4. International Research Visitors

### *7.4.1. Visits of International Scientists*

Victor Vianu, Professor at UC San Diego and holder of an Inria international chair, spent 3 months within Valda, employed as an ENS invited professor.

### *7.4.2. Visits to International Teams*

*7.4.2.1. Research Stays Abroad*

- Michaël Thomazo and Pierre Senellart have spent respectively two weeks and one week at TU Dresden, collaborating with Markus Krötzsch and Sebastian Rudolph.
- Pierre Senellart has spent a cumulated time of around three weeks at National University of Singapore, co-advising Debabrota Basu, PhD student working under the co-supervision of Stéphane Bressan, visiting Stéphane Bressan and other researchers at NUS, and participating in the French–Singapore workshop on AI, where Olivier Cappé represented CNRS.

# 8. Dissemination

## 8.1. Promoting Scientific Activities

### *8.1.1. Scientific Events Organisation*

*8.1.1.1. General Chair, Scientific Chair*

- Luc Segoufin, chair of the steering committee of the conference series *Highlights of Logic, Games and Automata*

*8.1.1.2. Member of the Organizing Committees*

- Luc Segoufin and Pierre Senellart, co-organizers of École de Printemps en Informatique Théorique (EPIT) 2019
- Pierre Senellart, co-organizer of ACM-ICPC Southwestern Europe 2018 competition

### *8.1.2. Scientific Events Selection*

*8.1.2.1. Chair of Conference Program Committees*

- Pierre Senellart, RoD (Reasoning on Data) workshop at The Web Conference 2018 (co-chair)

*8.1.2.2. Member of the Conference Program Committees*

- Camille Bourgaux, AAAI 2019
- Olivier Cappé, COLT 2018, ALT 2019
- Pierre Senellart, BDA 2018, PODS 2019
- Michaël Thomazo, IJCAI 2018, AAAI 2019

### *8.1.3. Journal*

*8.1.3.1. Member of the Editorial Boards*

- Olivier Cappé, associate editor, *Annals of the Institute of Statistical Mathematics*

*8.1.3.2. Reviewer - Reviewing Activities*

- Pierre Senellart, *Transactions on Database Systems*, *VLDB Journal*

### *8.1.4. Invited Talks*

- Pierre Senellart, keynote at Theory and Practice of Provenance (TaPP), London, United Kingdom
- Pierre Senellart, keynote at TempWeb workshop, The Web Conference, Lyon, France
- Pierre Senellart, LORIA Colloquium, Nancy, France

### *8.1.5. Leadership within the Scientific Community*

- Serge Abiteboul is a member of the French Academy of Sciences, of the Academia Europa, and of the scientific council of the Société Informatique de France.
- Pierre Senellart is a member of the steering committee of BDA, the French scientific community on data management.

### *8.1.6. Scientific Expertise*

- Pierre Senellart, FWO

### *8.1.7. Research Administration*

- Olivier Cappé is a scientific deputy director of CNRS division of Information Sciences and Technologies (INS2I).
- Luc Segoufin is a member of the CNHSCT of Inria.
- Pierre Senellart is a member of the board of section 6 of the National Committee for Scientific Research.
- Pierre Senellart is deputy director of the DI ENS laboratory, joint between ENS, CNRS, and Inria.
- Pierre Senellart is a member of the board of the DIM RFSI (Réseau Francilien en Sciences Informatiques).
- Pierre Senellart is a member of the scientific council of PGMO (Programme Gaspard Monge).

## 8.2. Teaching - Supervision - Juries

### *8.2.1. Teaching*

Licence: Pierre Senellart, *Databases*, 32 heqTD, L3, École normale supérieure
Licence: Pierre Senellart, *Algorithms*, 18 heqTD, L3, École normale supérieure
Licence: Michaël Thomazo, *Formal languages*, 22 heqTD, L3, Télécom ParisTech
Master: Serge Abiteboul & Pierre Senellart, *Web data management*, 36 heqTD, M2, MPRI

Pierre Senellart has various teaching responsibilities (L3 internships, M2 administration, tutoring, entrance competition) at ENS.

### *8.2.2. Supervision*

PhD: Miyoung Han, *Reinforcement learning approaches in dynamic environments*, Télécom ParisTech, 19 July 2018, Pierre Senellart

PhD: Alexandre Vigny, *Query enumeration and nowhere dense graphs*, Université Paris-Diderot, 27 September 2018, Arnaud Durand & Luc Segoufin

PhD: Mikaël Monet, *Combined complexity of probabilistic query evaluation*, Université Paris-Saclay, 12 October 2018, Antoine Amarilli & Pierre Senellart

PhD: Quentin Lobbé, *Archives, fragments Web et diasporas. Pour une exploration désagrégée de corpus d'archives Web liées aux représentations en ligne des diasporas.* Université Paris-Saclay, 9 November 2018, Dana Diminescu & Pierre Senellart

PhD in progress: Julien Grange, *Graph properties: order and arithmetic in predicate logics*, started in September 2017, Luc Segoufin

PhD in progress: Karima Rafes, *Le Linked Data à l'université : la plateforme LinkedWiki*, defense planned in January 2019, Serge Abiteboul & Sarah Cohen-Boulakia

PhD in progress: Yann Ramusat, *Provenance-based routing in probabilistic graphs*, started in September 2018, Silviu Maniu & Pierre Senellart

PhD in progess: Yoan Russac, *Sequential methods for robust decision making*, started in December 2018, Olivier Cappé

### *8.2.3. Juries*

- PhD Frederik Harwarth, June 2018, Humbold University, Luc Segoufin (reviewer)
- PhD Debabrota Basu, October 2018, National University of Singapore, Olivier Cappé (reviewer)
- PhD Govind, December 2018, Université Caen–Normandie, Pierre Senellart (president)
- PhD Ngurah Agus Sanjaya Er, December 2018, Télécom ParisTech, Pierre Senellart (president)

## 8.3. Popularization

### *8.3.1. Internal or external Inria responsibilities*

Serge Abiteboul is the president of the strategic committee of the Blaise Pascal foundation for scientific mediation.

### *8.3.2. Articles and contents*

Serge Abiteboul published *Le bot qui murmurait à l'oreille de la vieille dame* at the *Le Pommier* éditions, a collection of short stories on the digital world, accompanied with scientific and technical discussions.

Serge Abiteboul writes regular columns on popularization of computer science in La Recherche and Le Monde (Économie).

### *8.3.3. Education*

Pierre Senellart participated to a week-long meeting of teachers in *classes préparatoires* in Lumini in May 2018 to discuss the future of computer science education and to give an introduction to database research.

# 9. 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] A. AMARILLI, M. MONET, P. SENELLART.*Connecting Width and Structure in Knowledge Compilation*, in "ICDT 2018 - 21st International Conference on Database Theory", Vienna, Austria, Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, March 2018, vol. 98, p. 1-17 [*DOI :* 10.4230/LIPICS.ICDT.2018.6], https://hal.inria.fr/hal-01851564

[3] C. BOURGAUX, A. TURHAN.*Temporal Query Answering in DL-Lite over Inconsistent Data*, in "The Semantic Web - ISWC 2017 - 16th International Semantic Web Conference, Vienna, Austria, October 21-25, 2017, Proceedings, Part I", 2017, p. 121–137, https://doi.org/10.1007/978-3-319-68288-4_8

[4] F. JACQUEMARD, L. SEGOUFIN, J. DIMINO.*FO2(<, +1, ~) on data trees, data tree automata and branching vector addition systems*, in "Logical Methods in Computer Science", 2016, vol. 12, n° 2, https://doi.org/10.2168/LMCS-12(2:3)2016

[5] P. LAGRÉE, O. CAPPÉ, B. CAUTIS, S. MANIU.*Effective Large-Scale Online Influence Maximization*, in "2017 IEEE International Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, November 18-21, 2017", 2017, p. 937–942, https://doi.org/10.1109/ICDM.2017.118

[6] M. LECLÈRE, M.-L. MUGNIER, M. THOMAZO, F. ULLIANA.*A Single Approach to Decide Chase Termination on Linear Existential Rules*, in "DL 2018 - Description Logics", Tempe, United States, October 2018, https://arxiv.org/abs/1810.02132 , https://hal-lirmm.ccsd.cnrs.fr/lirmm-01892353

[7] S. MANIU, R. CHENG, P. SENELLART.*An Indexing Framework for Queries on Probabilistic Graphs*, in "ACM Trans. Datab. Syst", 2017, https://hal.inria.fr/hal-01437580

[8] D. MONTOYA, S. ABITEBOUL, P. SENELLART.*Hup-me: inferring and reconciling a timeline of user activity from rich smartphone data*, in "Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Bellevue, WA, USA, November 3-6, 2015", 2015, p. 62:1–62:4, http://doi.acm.org/10.1145/2820783.2820852

[9] N. SCHWEIKARDT, L. SEGOUFIN, A. VIGNY.*Enumeration for FO Queries over Nowhere Dense Graphs*, in "PODS 2018 - Principles Of Database Systems", Houston, United States, June 2018, https://hal.inria.fr/hal-01895786

[10] P. SENELLART, L. JACHIET, S. MANIU, Y. RAMUSAT.*ProvSQL: Provenance and Probability Management in PostgreSQL*, in "Proceedings of the VLDB Endowment (PVLDB)", August 2018, vol. 11, n° 12, p. 2034-2037 [*DOI :* 10.14778/3229863.3236253], https://hal.inria.fr/hal-01851538

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[11] M. HAN.*Reinforcement Learning Approaches in Dynamic Environments*, Télécom ParisTech, July 2018, https://hal.inria.fr/tel-01891805

[12] Q. LOBBÉ.*Archives, Web fragments and diasporas. For a disaggregated exploration of web archives related to online representations of diasporas*, Université Paris-Saclay, November 2018, https://hal.inria.fr/tel-01963548

[13] M. MONET.*Combined Complexity of Probabilistic Query Evaluation*, Université Paris-Saclay, October 2018, https://hal.inria.fr/tel-01963559

[14] A. VIGNY.*Query enumeration and nowhere dense graphs*, Université Paris-Diderot, September 2018, https://hal.inria.fr/tel-01963540

### Articles in International Peer-Reviewed Journal

[15] E. PITOURA, P. TSAPARAS, G. FLOURIS, I. FUNDULAKI, P. PAPADAKOS, S. ABITEBOUL, G. WEIKUM.*On Measuring Bias in Online Information*, in "SIGMOD record", 2018, p. 1-6, https://hal.inria.fr/hal-01638069

[16] P. SENELLART, L. JACHIET, S. MANIU, Y. RAMUSAT.*ProvSQL: Provenance and Probability Management in PostgreSQL*, in "Proceedings of the VLDB Endowment (PVLDB)", August 2018, vol. 11, n$^{o}$ 12, p. 2034-2037 [*DOI :* 10.14778/3229863.3236253], https://hal.inria.fr/hal-01851538

### Articles in National Peer-Reviewed Journal

[17] S. ABITEBOUL.*Les algorithmes du commerce*, in "Concurrences - revue des droits de la concurrence", 2018, https://hal.inria.fr/hal-01744283

[18] S. ABITEBOUL.*Les déclinaisons de la neutralité*, in "ANNALES DES MINES", December 2018, https://hal.inria.fr/hal-01963510

### International Conferences with Proceedings

[19] 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

[20] A. AMARILLI, M. MONET, P. SENELLART.*Connecting Width and Structure in Knowledge Compilation*, in "ICDT 2018 - 21st International Conference on Database Theory", Vienna, Austria, Leibniz International Proceedings in Informatics (LIPIcs), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, March 2018, vol. 98, p. 1-17 [*DOI :* 10.4230/LIPICS.ICDT.2018.6], https://hal.inria.fr/hal-01851564

[21] M. HAN, P.-H. WUILLEMIN, P. SENELLART.*Focused Crawling through Reinforcement Learning*, in "18th International Conference on Web Engineering (ICWE 2018)", Cáceres, Spain, T. MIKKONE, R. KLAMMA, J. HERNÁNDEZ (editors), Lecture Notes in Computer Science, Springer, June 2018, vol. 10845, p. 261-278 [*DOI :* 10.1007/978-3-319-91662-0_20], https://hal.inria.fr/hal-01851547

[22] M. LECLÈRE, M.-L. MUGNIER, M. THOMAZO, F. ULLIANA.*A Single Approach to Decide Chase Termination on Linear Existential Rules*, in "DL 2018 - Description Logics", Tempe, United States, October 2018, https://arxiv.org/abs/1810.02132 , https://hal-lirmm.ccsd.cnrs.fr/lirmm-01892353

[23] Q. LOBBÉ.*Revealing Historical Events out of Web Archives*, in "22nd International Conference on Theory and Practice of Digital Libraries (TPDL 2018)", Porto, Portugal, September 2018, https://hal.archives-ouvertes.fr/hal-01895951

[24] Q. LOBBÉ.*Where the dead blogs are. A Disaggregated Exploration of Web Archives to Reveal Extinct Online Collectives*, in "ICADL 2018 - 20th International Conference on Asia-Pacific Digital Libraries", Hamilton, New Zealand, November 2018, p. 1-12, https://hal.archives-ouvertes.fr/hal-01895955

[25] D. MONTOYA, T. P. TANON, S. ABITEBOUL, P. SENELLART, F. M. SUCHANEK.*A Knowledge Base for Personal Information Management*, in "LDOW2018 - Linked Open Data Workshop at the World Wide Web Conference", Lyon, France, April 2018, https://hal-imt.archives-ouvertes.fr/hal-01719312

[26] M. NIEWERTH, L. SEGOUFIN.*Enumeration of MSO Queries on Strings with Constant Delay and Logarithmic Updates*, in "Principles of Databse Systems, PODS'18", Houston, United States, ACM Press, June 2018 [*DOI :* 10.1145/3196959.3196961], https://hal.inria.fr/hal-01895796

[27] K. RAFES, S. ABITEBOUL, S. COHEN-BOULAKIA, B. RANCE.*Designing scientific SPARQL queries using autocompletion by snippets*, in "14th IEEE International Conference on eScience", Amsterdam, Netherlands, October 2018, https://hal.archives-ouvertes.fr/hal-01874780

[28] Y. RAMUSAT, S. MANIU, P. SENELLART.*Semiring Provenance over Graph Databases*, in "10th USENIX Workshop on the Theory and Practice of Provenance (TaPP 2018)", London, United Kingdom, July 2018, https://hal.inria.fr/hal-01850510

[29] N. SCHWEIKARDT, L. SEGOUFIN, A. VIGNY.*Enumeration for FO Queries over Nowhere Dense Graphs*, in "PODS 2018 - Principles Of Database Systems", Houston, United States, June 2018, https://hal.inria.fr/hal-01895786

[30] B. SPENCER, M. BENEDIKT, P. SENELLART.*Form Filling based on Constraint Solving*, in "18th International Conference on Web Engineering (ICWE 2018)", Cáceres, Spain, T. MIKKONE, R. KLAMMA, J. HERNÁNDEZ (editors), LNCS - Lecture Notes in Computer Science, Springer, June 2018, vol. 10845 [*DOI :* 10.1007/978-3-319-91662-0_20], https://hal.inria.fr/hal-01851555

### Scientific Books (or Scientific Book chapters)

[31] S. ABITEBOUL.*The Digital Shoebox*, in "Memory, edited by Philippe Tortell, Mark Turin, and Margot Young", UBC Press, October 2018, https://hal.inria.fr/hal-01875161

### Research Reports

[32] M. LECLÈRE, M.-L. MUGNIER, M. THOMAZO, F. ULLIANA.*A Single Approach to Decide Chase Termination on Linear Existential Rules*, arXiv:1810.02132, October 2018, https://hal-lirmm.ccsd.cnrs.fr/lirmm-01892375

### Other Publications

[33] A. AMARILLI, M. L. BA, D. DEUTCH, P. SENELLART.*Computing Possible and Certain Answers over Order-Incomplete Data*, October 2018, https://arxiv.org/abs/1801.06396 - 63 pages, 48 references. Submitted. Extended journal version of arXiv:1707.07222, https://hal.inria.fr/hal-01891811

[34] 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

[35] A. AMARILLI, M. MONET, P. SENELLART.*Connecting Width and Structure in Knowledge Compilation (Extended Version)*, May 2018, https://arxiv.org/abs/1709.06188 - 33 pages, no figures, 40 references. This is the full version with proofs of the corresponding ICDT'18 publication, and it integrates all reviewer

feedback. Except for the additional appendices, and except for formatting differences and inessential changes, the contents are the same as in the conference version [*DOI :* 10.4230/LIPIcs.ICDT.2018.6], https://hal.inria.fr/hal-01614551

[36] D. Basu, P. Senellart, S. Bressan.*BelMan: Bayesian Bandits on the Belief–Reward Manifold*, October 2018, https://arxiv.org/abs/1805.01627 - working paper or preprint, https://hal.inria.fr/hal-01891813

[37] W. Kazana, L. Segoufin.*First-order queries on classes of structures with bounded expansion*, February 2018, working paper or preprint, https://hal.inria.fr/hal-01706665

## References in notes

[38] S. Abiteboul, B. André, D. Kaplan.*Managing your digital life*, in "Commun. ACM", 2015, vol. 58, n$^{o}$ 5, p. 32–35, http://doi.acm.org/10.1145/2670528

[39] S. Abiteboul, P. Bourhis, V. Vianu.*Comparing workflow specification languages: A matter of views*, in "ACM Trans. Database Syst.", 2012, vol. 37, n$^{o}$ 2, p. 10:1–10:59, http://doi.acm.org/10.1145/2188349.2188352

[40] S. Abiteboul, P. Buneman, D. Suciu.*Data on the Web: From Relations to Semistructured Data and XML*, Morgan Kaufmann, 1999

[41] S. Abiteboul, D. Deutch, V. Vianu.*Deduction with Contradictions in Datalog*, in "Proc. 17th International Conference on Database Theory (ICDT), Athens, Greece, March 24-28, 2014", N. Schweikardt, V. Christophides, V. Leroy (editors), OpenProceedings.org, 2014, p. 143–154, https://doi.org/10.5441/002/icdt.2014.17

[42] S. Abiteboul, L. Herr, J. Van den Bussche.*Temporal Versus First-Order Logic to Query Temporal Databases*, in "Proceedings of the Fifteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 3-5, 1996, Montreal, Canada", R. Hull (editor), ACM Press, 1996, p. 49–57, http://doi.acm.org/10.1145/237661.237674

[43] S. Abiteboul, R. Hull, V. Vianu.*Foundations of Databases*, Addison-Wesley, 1995, http://webdam.inria.fr/Alice/

[44] S. Abiteboul, B. Kimelfeld, Y. Sagiv, P. Senellart.*On the expressiveness of probabilistic XML models*, in "VLDB J.", 2009, vol. 18, n$^{o}$ 5, p. 1041–1064, https://doi.org/10.1007/s00778-009-0146-1

[45] S. Abiteboul, I. Manolescu, P. Rigaux, M. Rousset, P. Senellart.*Web Data Management*, Cambridge University Press, 2011, http://webdam.inria.fr/Jorge

[46] S. Abiteboul, L. Segoufin, V. Vianu.*Representing and querying XML with incomplete information*, in "ACM Trans. Database Syst.", 2006, vol. 31, n$^{o}$ 1, p. 208–254, http://doi.acm.org/10.1145/1132863.1132869

[47] A. Amarilli, P. Bourhis, P. Senellart.*Provenance Circuits for Trees and Treelike Instances*, in "Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part II", 2015, p. 56–68, https://doi.org/10.1007/978-3-662-47666-6_5

[48] A. AMARILLI, P. BOURHIS, P. SENELLART.*Tractable Lineages on Treelike Instances: Limits and Extensions*, in "Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016", T. MILO, W. TAN (editors), ACM, 2016, p. 355–370, http://doi.acm.org/10.1145/2902251.2902301

[49] Y. AMSTERDAMER, D. DEUTCH, V. TANNEN.*Provenance for aggregate queries*, in "Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2011, June 12-16, 2011, Athens, Greece", M. LENZERINI, T. SCHWENTICK (editors), ACM, 2011, p. 153–164, http://doi.acm.org/10.1145/1989284.1989302

[50] Y. AMSTERDAMER, Y. GROSSMAN, T. MILO, P. SENELLART.*CrowdMiner: Mining association rules from the crowd*, in "PVLDB", 2013, vol. 6, n$^{o}$ 12, p. 1250–1253, http://www.vldb.org/pvldb/vol6/p1250-amsterdamer.pdf

[51] P. B. BAEZA.*Querying graph databases*, in "Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA - June 22 - 27, 2013", R. HULL, W. FAN (editors), ACM, 2013, p. 175–188, http://doi.acm.org/10.1145/2463664.2465216

[52] D. BARBARÁ, H. GARCIA-MOLINA, D. PORTER.*The Management of Probabilistic Data*, in "IEEE Trans. Knowl. Data Eng.", 1992, vol. 4, n$^{o}$ 5, p. 487–502, https://doi.org/10.1109/69.166990

[53] D. BASU, Q. LIN, W. CHEN, H. T. VO, Z. YUAN, P. SENELLART, S. BRESSAN.*Regularized Cost-Model Oblivious Database Tuning with Reinforcement Learning*, in "T. Large-Scale Data- and Knowledge-Centered Systems", 2016, vol. 28, p. 96–132, https://doi.org/10.1007/978-3-662-53455-7_5

[54] M. BENEDIKT, G. GOTTLOB, P. SENELLART.*Determining relevance of accesses at runtime*, in "Proceedings of the 30th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2011, June 12-16, 2011, Athens, Greece", M. LENZERINI, T. SCHWENTICK (editors), ACM, 2011, p. 211–222, http://doi.acm.org/10.1145/1989284.1989309

[55] M. BENEDIKT, P. SENELLART.*Databases*, in "Computer Science, The Hardware, Software and Heart of It", Springer, 2011, p. 169–229, https://doi.org/10.1007/978-1-4614-1168-0_10

[56] M. BIENVENU, D. DEUTCH, D. MARTINENGHI, P. SENELLART, F. M. SUCHANEK.*Dealing with the Deep Web and all its Quirks*, in "Proceedings of the Second International Workshop on Searching and Integrating New Web Data Sources, Istanbul, Turkey, August 31, 2012", M. BRAMBILLA, S. CERI, T. FURCHE, G. GOTTLOB (editors), CEUR Workshop Proceedings, CEUR-WS.org, 2012, vol. 884, p. 21–24, http://ceur-ws.org/Vol-884/VLDS2012_p21_Bienvenu.pdf

[57] M. BOJAŃCZYK, L. SEGOUFIN, S. TORUŃCZYK.*Verification of database-driven systems via amalgamation*, in "Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA - June 22 - 27, 2013", R. HULL, W. FAN (editors), ACM, 2013, p. 63–74, http://doi.acm.org/10.1145/2463664.2465228

[58] P. BUNEMAN, S. KHANNA, W.-C. TAN.*Why and Where: A Characterization of Data Provenance*, in "Database Theory - ICDT 2001, 8th International Conference, London, UK, January 4-6, 2001, Proceedings", J. VAN DEN BUSSCHE, V. VIANU (editors), Lecture Notes in Computer Science, Springer, 2001, vol. 1973, p. 316–330, https://doi.org/10.1007/3-540-44503-X_20

[59] B. COURCELLE.*The Monadic Second-Order Logic of Graphs. I. Recognizable Sets of Finite Graphs*, in "Inf. Comput.", 1990, vol. 85, n$^{o}$ 1, p. 12–75, https://doi.org/10.1016/0890-5401(90)90043-H

[60] N. N. DALVI, D. SUCIU.*The dichotomy of probabilistic inference for unions of conjunctive queries*, in "J. ACM", 2012, vol. 59, n$^{o}$ 6, p. 30:1–30:87, http://doi.acm.org/10.1145/2395116.2395119

[61] A. DESHPANDE, Z. G. IVES, V. RAMAN.*Adaptive Query Processing*, in "Foundations and Trends in Databases", 2007, vol. 1, n$^{o}$ 1, p. 1–140, https://doi.org/10.1561/1900000001

[62] P. DONMEZ, J. G. CARBONELL.*Proactive learning: cost-sensitive active learning with multiple imperfect oracles*, in "Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008, Napa Valley, California, USA, October 26-30, 2008", J. G. SHANAHAN, S. AMER-YAHIA, I. MANOLESCU, Y. ZHANG, D. A. EVANS, A. KOLCZ, K. CHOI, A. CHOWDHURY (editors), ACM, 2008, p. 619–628, http://doi.acm.org/10.1145/1458082.1458165

[63] M. FAHEEM, P. SENELLART.*Adaptive Web Crawling Through Structure-Based Link Classification*, in "Digital Libraries: Providing Quality Information - 17th International Conference on Asia-Pacific Digital Libraries, ICADL 2015, Seoul, Korea, December 9-12, 2015, Proceedings", R. B. ALLEN, J. HUNTER, M. L. ZENG (editors), Lecture Notes in Computer Science, Springer, 2015, vol. 9469, p. 39–51, https://doi.org/10.1007/978-3-319-27974-9_5

[64] N. FRANCIS, L. SEGOUFIN, C. SIRANGELO.*Datalog Rewritings of Regular Path Queries using Views*, in "Logical Methods in Computer Science", 2015, vol. 11, n$^{o}$ 4, https://doi.org/10.2168/LMCS-11(4:14)2015

[65] A. GALLAND, S. ABITEBOUL, A. MARIAN, P. SENELLART.*Corroborating information from disagreeing views*, in "Proceedings of the Third International Conference on Web Search and Web Data Mining, WSDM 2010, New York, NY, USA, February 4-6, 2010", B. D. DAVISON, T. SUEL, N. CRASWELL, B. LIU (editors), ACM, 2010, p. 131–140, http://doi.acm.org/10.1145/1718487.1718504

[66] F. GEERTS, A. POGGI.*On database query languages for K-relations*, in "J. Applied Logic", 2010, vol. 8, n$^{o}$ 2, p. 173–185, https://doi.org/10.1016/j.jal.2009.09.001

[67] L. GETOOR.*Introduction to statistical relational learning*, MIT Press, 2007

[68] G. GOURITEN, S. MANIU, P. SENELLART.*Scalable, generic, and adaptive systems for focused crawling*, in "25th ACM Conference on Hypertext and Social Media, HT '14, Santiago, Chile, September 1-4, 2014", L. FERRES, G. ROSSI, V. A. F. ALMEIDA, E. HERDER (editors), ACM, 2014, p. 35–45, http://doi.acm.org/10.1145/2631775.2631795

[69] T. J. GREEN, G. KARVOUNARAKIS, V. TANNEN.*Provenance semirings*, in "Proceedings of the Twenty-Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 11-13, 2007, Beijing, China", L. LIBKIN (editor), ACM, 2007, p. 31–40, http://doi.acm.org/10.1145/1265530.1265535

[70] T. J. GREEN, V. TANNEN.*Models for Incomplete and Probabilistic Information*, in "IEEE Data Eng. Bull.", 2006, vol. 29, n$^{o}$ 1, p. 17–24, http://sites.computer.org/debull/A06mar/green.ps

[71] A. Y. HALEVY.*Answering queries using views: A survey*, in "VLDB J.", 2001, vol. 10, n$^{o}$ 4, p. 270–294, https://doi.org/10.1007/s007780100054

[72] M. A. HEARST, S. T. DUMAIS, E. OSUNA, J. PLATT, B. SCHOLKOPF.*Support vector machines*, in "IEEE Intelligent Systems", 1998, vol. 13, n$^{o}$ 4, p. 18–28, https://doi.org/10.1109/5254.708428

[73] T. IMIELINSKI, W. LIPSKI JR..*Incomplete Information in Relational Databases*, in "J. ACM", 1984, vol. 31, n$^{o}$ 4, p. 761–791, http://doi.acm.org/10.1145/1634.1886

[74] W. KAZANA, L. SEGOUFIN.*Enumeration of first-order queries on classes of structures with bounded expansion*, in "Proceedings of the 32nd ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2013, New York, NY, USA - June 22 - 27, 2013", R. HULL, W. FAN (editors), ACM, 2013, p. 297–308, http://doi.acm.org/10.1145/2463664.2463667

[75] B. KIMELFELD, P. SENELLART.*Probabilistic XML: Models and Complexity*, in "Advances in Probabilistic Databases for Uncertain Information Management", Z. MA, L. YAN (editors), Studies in Fuzziness and Soft Computing, Springer, 2013, vol. 304, p. 39–66, https://doi.org/10.1007/978-3-642-37509-5_3

[76] A. C. KLUG.*Equivalence of Relational Algebra and Relational Calculus Query Languages Having Aggregate Functions*, in "J. ACM", 1982, vol. 29, n$^{o}$ 3, p. 699–717, http://doi.acm.org/10.1145/322326.322332

[77] D. KOSSMANN.*The State of the art in distributed query processing*, in "ACM Comput. Surv.", 2000, vol. 32, n$^{o}$ 4, p. 422–469, http://doi.acm.org/10.1145/371578.371598

[78] J. D. LAFFERTY, A. MCCALLUM, F. C. N. PEREIRA.*Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data*, in "Proceedings of the Eighteenth International Conference on Machine Learning (ICML 2001), Williams College, Williamstown, MA, USA, June 28 - July 1, 2001", C. E. BRODLEY, A. P. DANYLUK (editors), Morgan Kaufmann, 2001, p. 282–289

[79] S. LEI, S. MANIU, L. MO, R. CHENG, P. SENELLART.*Online Influence Maximization*, in "Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015", 2015, p. 645–654, http://doi.acm.org/10.1145/2783258.2783271

[80] M. MOHRI.*Semiring Frameworks and Algorithms for Shortest-Distance Problems*, in "Journal of Automata, Languages and Combinatorics", 2002, vol. 7, n$^{o}$ 3, p. 321–350

[81] F. NEVEN.*Automata Theory for XML Researchers*, in "SIGMOD Record", 2002, vol. 31, n$^{o}$ 3, p. 39–46, http://doi.acm.org/10.1145/601858.601869

[82] L. SEGOUFIN.*A glimpse on constant delay enumeration (Invited Talk)*, in "31st International Symposium on Theoretical Aspects of Computer Science (STACS 2014), STACS 2014, March 5-8, 2014, Lyon, France", E. W. MAYR, N. PORTIER (editors), LIPIcs, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2014, vol. 25, p. 13–27, https://doi.org/10.4230/LIPIcs.STACS.2014.13

[83] P. SENELLART, A. MITTAL, D. MUSCHICK, R. GILLERON, M. TOMMASI.*Automatic wrapper induction from hidden-web sources with domain knowledge*, in "10th ACM International Workshop on Web Information and Data Management (WIDM 2008), Napa Valley, California, USA, October 30, 2008", C. Y. CHAN, N. POLYZOTIS (editors), ACM, 2008, p. 9–16, http://doi.acm.org/10.1145/1458502.1458505

[84] B. SETTLES, M. CRAVEN, L. FRIEDLAND.*Active learning with real annotation costs*, in "NIPS 2008 Workshop on Cost-Sensitive Learning", 2008, http://burrsettles.com/pub/settles.nips08ws.pdf

[85] B. SETTLES.*Active Learning*, Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan & Claypool Publishers, 2012, https://doi.org/10.2200/S00429ED1V01Y201207AIM018

[86] F. M. SUCHANEK, S. ABITEBOUL, P. SENELLART.*PARIS: Probabilistic Alignment of Relations, Instances, and Schema*, in "PVLDB", 2011, vol. 5, n$^{o}$ 3, p. 157–168, http://www.vldb.org/pvldb/vol5/p157_fabianmsuchanek_vldb2012.pdf

[87] D. SUCIU, D. OLTEANU, C. RÉ, C. KOCH.*Probabilistic Databases*, Synthesis Lectures on Data Management, Morgan & Claypool Publishers, 2011, https://doi.org/10.2200/S00362ED1V01Y201105DTM016

[88] R. S. SUTTON, A. G. BARTO.*Reinforcement learning - an introduction*, Adaptive computation and machine learning, MIT Press, 1998, http://www.worldcat.org/oclc/37293240

[89] M. Y. VARDI.*The Complexity of Relational Query Languages (Extended Abstract)*, in "Proceedings of the 14th Annual ACM Symposium on Theory of Computing, May 5-7, 1982, San Francisco, California, USA", H. R. LEWIS, B. B. SIMONS, W. A. BURKHARD, L. H. LANDWEBER (editors), ACM, 1982, p. 137–146, http://doi.acm.org/10.1145/800070.802186

[90] K. ZHOU, M. LALMAS, T. SAKAI, R. CUMMINS, J. M. JOSE.*On the reliability and intuitiveness of aggregated search metrics*, in "22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013", Q. HE, A. IYENGAR, W. NEJDL, J. PEI, R. RASTOGI (editors), ACM, 2013, p. 689–698, http://doi.acm.org/10.1145/2505515.2505691

[91] M. T. ÖZSU, P. VALDURIEZ.*Principles of Distributed Database Systems, Third Edition*, Springer, 2011, https://doi.org/10.1007/978-1-4419-8834-8

# Project-Team WHISPER

## Well Honed Infrastructure Software for Programming Environments and Runtimes

IN COLLABORATION WITH: Laboratoire d'informatique de Paris 6 (LIP6)

IN PARTNERSHIP WITH:
**CNRS**

**Sorbonne Université (UPMC)**

RESEARCH CENTER
**Paris**

THEME
**Distributed Systems and middleware**

# Table of contents

## Project-Team WHISPER

*Creation of the Team: 2014 May 15, updated into Project-Team: 2015 December 01*

**Keywords:**

**Computer Science and Digital Science:**

A1. - Architectures, systems and networks
A1.1.1. - Multicore, Manycore
A2.1.6. - Concurrent programming
A2.1.10. - Domain-specific languages
A2.1.11. - Proof languages
A2.2.1. - Static analysis
A2.2.5. - Run-time systems
A2.3.1. - Embedded systems
A2.3.3. - Real-time systems
A2.4. - Formal method for verification, reliability, certification
A2.4.3. - Proofs
A2.5. - Software engineering
A2.6.1. - Operating systems
A2.6.2. - Middleware
A2.6.3. - Virtual machines

**Other Research Topics and Application Domains:**

B5. - Industry of the future
B5.2.1. - Road vehicles
B5.2.3. - Aviation
B5.2.4. - Aerospace
B6.1. - Software industry
B6.1.1. - Software engineering
B6.1.2. - Software evolution, maintenance
B6.3.3. - Network Management
B6.5. - Information systems
B6.6. - Embedded systems

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Pierre-Évariste Dagand [CNRS, Researcher]
Julia Lawall [Inria, Senior Researcher]
Gilles Muller [Team Leader, Inria, Senior Researcher, HDR]

**Faculty Member**

Bertil Folliot [Univ Pierre et Marie Curie, Professor, HDR]

**Technical Staff**

Antoine Blin [Inria, until Oct 2018, granted by ORANGE SA]

**PhD Students**

Cédric Courtaud [Thales]
Redha Gouicem [Univ Pierre et Marie Curie]
Darius Mercadier [Univ Pierre et Marie Curie]
Lucas Serrano [Univ Pierre et Marie Curie]

**Post-Doctoral Fellow**
Van-Anh Nguyen [Univ Pierre et Marie Curie, financed by ANR ITrans]

**Administrative Assistants**
Nelly Maloisel [Inria]
Eugène Kamdem [UPMC, Assistant]

# 2. Overall Objectives

## 2.1. Overall Objectives

The focus of Whisper is on how to develop (new) and improve (existing) infrastructure software. Infrastructure software (also called systems software) is the software that underlies all computing. Such software allows applications to access resources and provides essential services such as memory management, synchronization and inter-process interactions. Starting bottom-up from the hardware, examples include virtual machine hypervisors, operating systems, managed runtime environments, standard libraries, and browsers, which amount to the new operating system layer for Internet applications. For such software, efficiency and correctness are fundamental. Any overhead will impact the performance of all supported applications. Any failure will prevent the supported applications from running correctly. Since computing now pervades our society, with few paper backup solutions, correctness of software at all levels is critical. Formal methods are increasingly being applied to operating systems code in the research community [37], [42], [80]. Still, such efforts require a huge amount of manpower and a high degree of expertise which makes this work difficult to replicate in standard infrastructure-software development.

In terms of methodology, Whisper is at the interface of the domains of operating systems, software engineering and programming languages. Our approach is to combine the study of problems in the development of real-world infrastructure software with concepts in programming language design and implementation, *e.g.*, of domain-specific languages, and knowledge of low-level system behavior. A focus of our work is on providing support for legacy code, while taking the needs and competences of ordinary system developers into account.

We aim at providing solutions that can be easily learned and adopted by system developers in the short term. Such solutions can be tools, such as Coccinelle [1], [8], [9] for transforming C programs, or domain-specific languages such as Devil [7] and Bossa [6] for designing drivers and kernel schedulers. Due to the small size of the team, Whisper mainly targets operating system kernels and runtimes for programming languages. We put an emphasis on achieving measurable improvements in performance and safety in practice, and on feeding these improvements back to the infrastructure software developer community.

# 3. Research Program

## 3.1. Scientific Foundations

### *3.1.1. Program analysis*

A fundamental goal of the research in the Whisper team is to elicit and exploit the knowledge found in existing code. To do this in a way that scales to a large code base, systematic methods are needed to infer code properties. We may build on either static [28], [30], [32] or dynamic analysis [51], [53], [59]. Static analysis consists of approximating the behavior of the source code from the source code alone, while dynamic analysis draws conclusions from observations of sample executions, typically of test cases. While dynamic

analysis can be more accurate, because it has access to information about actual program behavior, obtaining adequate test cases is difficult. This difficulty is compounded for infrastructure software, where many, often obscure, cases must be handled, and external effects such as timing can have a significant impact. Thus, we expect to primarily use static analyses. Static analyses come in a range of flavors, varying in the extent to which the analysis is *sound*, *i.e.*, the extent to which the results are guaranteed to reflect possible run-time behaviors.

One form of sound static analysis is *abstract interpretation* [30]. In abstract interpretation, atomic terms are interpreted as sound abstractions of their values, and operators are interpreted as functions that soundly manipulate these abstract values. The analysis is then performed by interpreting the program in a compositional manner using these abstracted values and operators. Alternatively, *dataflow analysis* [41] iteratively infers connections between variable definitions and uses, in terms of local transition rules that describe how various kinds of program constructs may impact variable values. Schmidt has explored the relationship between abstract interpretation and dataflow analysis [67]. More recently, more general forms of symbolic execution [28] have emerged as a means of understanding complex code. In symbolic execution, concrete values are used when available, and these are complemented by constraints that are inferred from terms for which only partial information is available. Reasoning about these constraints is then used to prune infeasible paths, and obtain more precise results. A number of works apply symbolic execution to operating systems code [25], [26].

While sound approaches are guaranteed to give correct results, they typically do not scale to the very diverse code bases that are prevalent in infrastructure software. An important insight of Engler et al. [35] was that valuable information could be obtained even when sacrificing soundness, and that sacrificing soundness could make it possible to treat software at the scales of the kernels of the Linux or BSD operating systems. Indeed, for certain types of problems, on certain code bases, that may mostly follow certain coding conventions, it may mostly be safe to *e.g.*, ignore the effects of aliases, assume that variable values are unchanged by calls to unanalyzed functions, etc. Real code has to be understood by developers and thus cannot be too complicated, so such simplifying assumptions are likely to hold in practice. Nevertheless, approaches that sacrifice soundness also require the user to manually validate the results. Still, it is likely to be much more efficient for the user to perform a potentially complex manual analysis in a specific case, rather than to implement all possible required analyses and apply them everywhere in the code base. A refinement of unsound analysis is the CEGAR approach [29], in which a highly approximate analysis is complemented by a sound analysis that checks the individual reports of the approximate analysis, and then any errors in reasoning detected by the sound analysis are used to refine the approximate analysis. The CEGAR approach has been applied effectively on device driver code in tools developed at Microsoft [17]. The environment in which the driver executes, however, is still represented by possibly unsound approximations.

Going further in the direction of sacrificing soundness for scalability, the software engineering community has recently explored a number of approaches to code understanding based on techniques developed in the areas of natural language understanding, data mining, and information retrieval. These approaches view code, as well as other software-reated artifacts, such as documentation and postings on mailing lists, as bags of words structured in various ways. Statistical methods are then used to collect words or phrases that seem to be highly correlated, independently of the semantics of the program constructs that connect them. The obliviousness to program semantics can lead to many false positives (invalid conclusions) [47], but can also highlight trends that are not apparent at the low level of individual program statements. We have previously explored combining such statistical methods with more traditional static analysis in identifying faults in the usage of constants in Linux kernel code [45].

### *3.1.2. Domain Specific Languages*

Writing low-level infrastructure code is tedious and difficult, and verifying it is even more so. To produce non-trivial programs, we could benefit from moving up the abstraction stack to enable both programming and proving as quickly as possible. Domain-specific languages (DSLs), also known as *little languages*, are a means to that end [5] [54].

*3.1.2.1. Traditional approach.*

Using little languages to aid in software development is a tried-and-trusted technique [70] by which programmers can express high-level ideas about the system at hand and avoid writing large quantities of formulaic C boilerplate.

This approach is typified by the Devil language for hardware access [7]. An OS programmer describes the register set of a hardware device in the high-level Devil language, which is then compiled into a library providing C functions to read and write values from the device registers. In doing so, Devil frees the programmer from having to write extensive bit-manipulation macros or inline functions to map between the values the OS code deals with, and the bit-representation used by the hardware: Devil generates code to do this automatically.

However, DSLs are not restricted to being "stub" compilers from declarative specifications. The Bossa language [6] is a prime example of a DSL involving imperative code (syntactically close to C) while offering a high-level of abstraction. This design of Bossa enables the developer to implement new process scheduling policies at a level of abstraction tailored to the application domain.

Conceptually, a DSL both abstracts away low-level details and justifies the abstraction by its semantics. In principle, it reduces development time by allowing the programmer to focus on high-level abstractions. The programmer needs to write less code, in a language with syntax and type checks adapted to the problem at hand, thus reducing the likelihood of errors.

*3.1.2.2. Embedding DSLs.*

The idea of a DSL has yet to realize its full potential in the OS community. Indeed, with the notable exception of interface definition languages for remote procedure call (RPC) stubs, most OS code is still written in a low-level language, such as C. Where DSL code generators are used in an OS, they tend to be extremely simple in both syntax and semantics. We conjecture that the effort to implement a given DSL usually outweighs its benefit. We identify several serious obstacles to using DSLs to build a modern OS: specifying what the generated code will look like, evolving the DSL over time, debugging generated code, implementing a bug-free code generator, and testing the DSL compiler.

Filet-o-Fish (FoF) [3] addresses these issues by providing a framework in which to build correct code generators from semantic specifications. This framework is presented as a Haskell library, enabling DSL writers to *embed* their languages within Haskell. DSL compilers built using FoF are quick to write, simple, and compact, but encode rigorous semantics for the generated code. They allow formal proofs of the run-time behavior of generated code, and automated testing of the code generator based on randomized inputs, providing greater test coverage than is usually feasible in a DSL. The use of FoF results in DSL compilers that OS developers can quickly implement and evolve, and that generate provably correct code. FoF has been used to build a number of domain-specific languages used in Barrelfish, [18] an OS for heterogeneous multicore systems developed at ETH Zurich.

The development of an embedded DSL requires a few supporting abstractions in the host programming language. FoF was developed in the purely functional language Haskell, thus benefiting from the type class mechanism for overloading, a flexible parser offering convenient syntactic sugar, and purity enabling a more algebraic approach based on small, composable combinators. Object-oriented languages – such as Smalltalk [36] and its descendant Pharo [22] – or multi-paradigm languages – such as the Scala programming language [56] – also offer a wide range of mechanisms enabling the development of embedded DSLs. Perhaps suprisingly, a low-level imperative language – such as C – can also be extended so as to enable the development of embedded compilers [19].

*3.1.2.3. Certifying DSLs.*

Whilst automated and interactive software verification tools are progressively being applied to larger and larger programs, we have not yet reached the point where large-scale, legacy software – such as the Linux kernel – could formally be proved "correct". DSLs enable a pragmatic approach, by which one could realistically strengthen a large legacy software by first narrowing down its critical component(s) and then focus our verification efforts onto these components.

Dependently-typed languages, such as Coq or Idris, offer an ideal environment for embedding DSLs [27], [23] in a unified framework enabling verification. Dependent types support the type-safe embedding of object languages and Coq's mixfix notation system enables reasonably idiomatic domain-specific concrete syntax. Coq's powerful abstraction facilities provide a flexible framework in which to not only implement and verify a range of domain-specific compilers [3], but also to combine them, and reason about their combination.

Working with many DSLs optimizes the "horizontal" compositionality of systems, and favors reuse of building blocks, by contrast with the "vertical" composition of the traditional compiler pipeline, involving a stack of comparatively large intermediate languages that are harder to reuse the higher one goes. The idea of building compilers from reusable building blocks is a common one, of course. But the interface contracts of such blocks tend to be complex, so combinations are hard to get right. We believe that being able to write and verify formal specifications for the pieces will make it possible to know when components can be combined, and should help in designing good interfaces.

Furthermore, the fact that Coq is also a system for formalizing mathematics enables one to establish a close, formal connection between embedded DSLs and non-trivial domain-specific models. The possibility of developing software in a truly "model-driven" way is an exciting one. Following this methodology, we have implemented a certified compiler from regular expressions to x86 machine code [4]. Interestingly, our development crucially relied on an existing Coq formalization, due to Braibant and Pous, [24] of the theory of Kleene algebras.

While these individual experiments seem to converge toward embedding domain-specific languages in rich type theories, further experimental validation is required. Indeed, Barrelfish is an extremely small software compared to the Linux kernel. The challenge lies in scaling this methodology up to large software systems. Doing so calls for a unified platform enabling the development of a myriad of DSLs, supporting code reuse across DSLs as well as providing support for mechanically-verified proofs.

## 3.2. Research direction: Tools for improving legacy infrastructure software

A cornerstone of our work on legacy infrastructure software is the Coccinelle program matching and transformation tool for C code. Coccinelle has been in continuous development since 2005. Today, Coccinelle is extensively used in the context of Linux kernel development, as well as in the development of other software, such as wine, python, kvm, and systemd. Currently, Coccinelle is a mature software project, and no research is being conducted on Coccinelle itself. Instead, we leverage Coccinelle in other research projects [20], [21], [57], [60], [64], [66], [68], [52], [46], both for code exploration, to better understand at a large scale problems in Linux development, and as an essential component in tools that require program matching and transformation. The continuing development and use of Coccinelle is also a source of visibility in the Linux kernel developer community. We submitted the first patches to the Linux kernel based on Coccinelle in 2007. Since then, over 5500 patches have been accepted into the Linux kernel based on the use of Coccinelle, including around 3000 by over 500 developers from outside our research group.

Our recent work has focused on driver porting. Specifically, we have considered the problem of porting a Linux device driver across versions, particularly backporting, in which a modern driver needs to be used by a client who, typically for reasons of stability, is not able to update their Linux kernel to the most recent version. When multiple drivers need to be backported, they typically need many common changes, suggesting that Coccinelle could be applicable. Using Coccinelle, however, requires writing backporting transformation rules. In order to more fully automate the backporting (or symmetrically forward porting) process, these rules should be generated automatically. We have carried out a preliminary study in this direction with David Lo of Singapore Management University; this work, published at ICSME 2016 [73], is limited to a port from one version to the next one, in the case where the amount of change required is limited to a single line of code. Whisper has been awarded an ANR PRCI grant to collaborate with the group of David Lo on scaling up the rule inference process and proposing a fully automatic porting solution.

## 3.3. Research direction: developing infrastructure software using Domain Specific Languages

We wish to pursue a *declarative* approach to developing infrastructure software. Indeed, there exists a significant gap between the high-level objectives of these systems and their implementation in low-level, imperative programming languages. To bridge that gap, we propose an approach based on domain-specific languages (DSLs). By abstracting away boilerplate code, DSLs increase the productivity of systems programmers. By providing a more declarative language, DSLs reduce the complexity of code, thus the likelihood of bugs.

Traditionally, systems are built by accretion of several, independent DSLs. For example, one might use Devil [7] to interact with devices, Bossa [6] to implement the scheduling policies. However, much effort is duplicated in implementing the back-ends of the individual DSLs. Our long term goal is to design a unified framework for developing and composing DSLs, following our work on Filet-o-Fish [3]. By providing a single conceptual framework, we hope to amortize the development cost of a myriad of DSLs through a principled approach to reusing and composing them.

Beyond the software engineering aspects, a unified platform brings us closer to the implementation of mechanically-verified DSLs. Using the Coq proof assistant as an x86 macro-assembler [4] is a step in that direction, which belongs to a larger trend of hosting DSLs in dependent type theories [23], [27], [55]. A key benefit of those approaches is to provide – by construction – a formal, mechanized semantics to the DSLs thus developed. This semantics offers a foundation on which to base further verification efforts, whilst allowing interaction with non-verified code. We advocate a methodology based on incremental, piece-wise verification. Whilst building fully-certified systems from the top-down is a worthwhile endeavor [42], we wish to explore a bottom-up approach by which one focuses first and foremost on crucial subsystems and their associated properties.

Our current work on DSLs has two complementary goals: (i) the design of a unified framework for developing and composing DSLs, following our work on Filet-o-Fish, and (ii) the design of domain-specific languages for domains where there is a critical need for code correctness, and corresponding methodologies for proving properties of the run-time behavior of the system.

# 4. Application Domains

## 4.1. Linux

Linux is an open-source operating system that is used in settings ranging from embedded systems to supercomputers. The most recent release of the Linux kernel, v4.14, comprises over 16 million lines of code, and supports 30 different families of CPU architectures, around 50 file systems, and thousands of device drivers. Linux is also in a rapid stage of development, with new versions being released roughly every 2.5 months. Recent versions have each incorporated around 13,500 commits, from around 1500 developers. These developers have a wide range of expertise, with some providing hundreds of patches per release, while others have contributed only one. Overall, the Linux kernel is critical software, but software in which the quality of the developed source code is highly variable. These features, combined with the fact that the Linux community is open to contributions and to the use of tools, make the Linux kernel an attractive target for software researchers. Tools that result from research can be directly integrated into the development of real software, where it can have a high, visible impact.

Starting from the work of Engler et al. [34], numerous research tools have been applied to the Linux kernel, typically for finding bugs [32], [50], [61], [72] or for computing software metrics [39], [78]. In our work, we have studied generic C bugs in Linux code [9], bugs in function protocol usage [43], [44], issues related to the processing of bug reports [65] and crash dumps [38], and the problem of backporting [60], [73], illustrating the variety of issues that can be explored on this code base. Unique among research groups working in this area, we have furthermore developed numerous contacts in the Linux developer community. These contacts provide insights into the problems actually faced by developers and serve as a means of validating the practical relevance of our work.

## 4.2. Device Drivers

Device drivers are essential to modern computing, to provide applications with access, via the operating system, to physical devices such as keyboards, disks, networks, and cameras. Development of new computing paradigms, such as the internet of things, is hampered because device driver development is challenging and error-prone, requiring a high level of expertise in both the targeted OS and the specific device. Furthermore, implementing just one driver is often not sufficient; today's computing landscape is characterized by a number of OSes, *e.g.*, Linux, Windows, MacOS, BSD and many real time OSes, and each is found in a wide range of variants and versions. All of these factors make the development, porting, backporting, and maintenance of device drivers a critical problem for device manufacturers, industry that requires specific devices, and even for ordinary users.

The last fifteen years have seen a number of approaches directed towards easing device driver development. Réveillère, who was supervised by G. Muller, proposes Devil [7], a domain-specific language for describing the low-level interface of a device. Chipounov *et al.* propose RevNic, [26] a template-based approach for porting device drivers from one OS to another. Ryzhyk *et al.* propose Termite, [62], [63] an approach for synthesizing device driver code from a specification of an OS and a device. Currently, these approaches have been successfully applied to only a small number of toy drivers. Indeed, Kadav and Swift [40] observe that these approaches make assumptions that are not satisfied by many drivers; for example, the assumption that a driver involves little computation other than the direct interaction between the OS and the device. At the same time, a number of tools have been developed for finding bugs in driver code. These tools include SDV [17], Coverity [34], CP-Miner, [49] PR-Miner [50], and Coccinelle [8]. These approaches, however, focus on analyzing existing code, and do not provide guidelines on structuring drivers.

In summary, there is still a need for a methodology that first helps the developer understand the software architecture of drivers for commonly used operating systems, and then provides tools for the maintenance of existing drivers.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

The Whisper team published three papers at USENIX ATC, one of the major conferences of our domain:

- Coccinelle: 10 Years of Automated Evolution in the Linux Kernel. J. Lawall and G.Muller. [14]
- DSAC: Effective Static Analysis of Sleep-in-Atomic-Context Bugs in Kernel Modules. J.-J. Bai, Y.-P. Wang, J. Lawall, S.-M. Hu. [12]
- The Battle of the Schedulers: FreeBSD ULE vs. Linux CFS. J. Bouron, S. Chevalley, B. Lepers, W. Zwaenepoel, R. Gouicem, J. Lawall, G. Muller, J. Sopena. [13]

Gilles Muller was co-PC chair of DSN 2018, the premier venue for dependable systems.

Julia Lawall was co-PC chair of the ASE 2018 Tool Demo track, in preparation for being the co-PC chair of the main ASE research paper track in 2019.

### *5.1.1. Awards*

The original work on Coccinelle "Documenting and automating collateral evolutions in Linux device drivers" [8] received an ACM EuroSys Test-of-Time award, recognizing it as the paper from EuroSys 2008 that is having the most lasting and current impact (http://eurosys2018.org/awards/).

# 6. New Software and Platforms

## 6.1. Coccinelle

KEYWORDS: Code quality - Evolution - Infrastructure software

FUNCTIONAL DESCRIPTION: Coccinelle is a tool for code search and transformation for C programs. It has been extensively used for bug finding and evolutions in Linux kernel code.

- Participants: Gilles Muller, Julia Lawall, Nicolas Palix, Rene Rydhof Hansen and Thierry Martinez
- Partners: LIP6 - IRILL
- Contact: Julia Lawall
- URL: http://coccinelle.lip6.fr

## 6.2. Prequel

KEYWORDS: Code search - Git
SCIENTIFIC DESCRIPTION: The commit history of a code base such as the Linux kernel is a gold mine of information on how evolutions should be made, how bugs should be fixed, etc. Nevertheless, the high volume of commits available and the rudimentary filtering tools provided mean that it is often necessary to wade through a lot of irrelevant information before finding example commits that can help with a specific software development problem. To address this issue, we propose Prequel (Patch Query Language), which brings the descriptive power of code matching to the problem of querying a commit history.
FUNCTIONAL DESCRIPTION: Prequel is a tool for searching for complex patterns in the commits of software managed using git.

- Participants: Gilles Muller and Julia Lawall
- Partners: LIP6 - IRILL
- Contact: Julia Lawall
- URL: http://prequel-pql.gforge.inria.fr/

## 6.3. Usuba

KEYWORDS: Cryptography - Optimizing compiler - Synchorous language
FUNCTIONAL DESCRIPTION: Usuba is a programming language for specifying block ciphers as well as a bitslicing compiler, for producing high-throughput and secure code.

- Contact: Pierre-Evariste Dagand
- Publication: Usuba, Optimizing & Trustworthy Bitslicing Compiler
- URL: https://github.com/DadaIsCrazy/usuba/

# 7. New Results

## 7.1. Software engineering for infrastructure software

The most visible tool developed in the Whisper team is Coccinelle, which this year marked the 10th anniversary of its release in open source. The paper "Coccinelle: 10 Years of Automated Evolution in the Linux Kernel," published at USENIX ATC'18 [14], traced the history of Coccinelle, its underlying design decisions and impact. The Coccinelle C-code matching and transformation tool was first released in 2008 to facilitate specification and automation in the evolution of Linux kernel code. The novel contribution of Coccinelle was to allow software developers to write code manipulation rules in terms of the code structure itself, via a generalization of the patch syntax. Over the years, Coccinelle has been extensively used in Linux kernel development, resulting in over 6000 commits to the Linux kernel, and has found its place as part of the Linux kernel development process. The USENIX ATC paper studies the impact of Coccinelle on Linux kernel development and the features of Coccinelle that have made it possible. It provides guidance on how other research-based tools can achieve practical impact in the open-source development community. This work was also presented to Linux kernel developers at Kernel Recipes and Open Source Summit Europe, and at the 8th Inria/Technicolor Workshop On Systems.

In a modern OS, kernel modules often use spinlocks and interrupt handlers to monopolize a CPU core to execute concurrent code in atomic context. In this situation, if the kernel module performs an operation that can sleep at runtime, a system hang may occur. We refer to this kind of concurrency bug as a sleep-in-atomic-context (SAC) bug. In practice, SAC bugs have received insufficient attention and are hard to find, as they do not always cause problems in real executions. In a paper published at USENIX ATC'18 [12], we propose a practical static approach named DSAC, to effectively detect SAC bugs and automatically recommend patches to help fix them. DSAC uses four key techniques: (1) a hybrid of flow-sensitive and -insensitive analysis to perform accurate and efficient code analysis; (2) a heuristics-based method to accurately extract kernel interfaces that can sleep at runtime; (3) a path-check method to effectively filter out repeated reports and false bugs; (4) a pattern-based method to automatically generate recommended patches to help fix the bugs. We evaluate DSAC on kernel modules (drivers, file systems, and network modules) of the Linux kernel, and on the FreeBSD and NetBSD kernels, and in total find 401 new real bugs. 272 of these bugs have been confirmed by the relevant kernel maintainers, and 43 patches generated by DSAC have been applied by kernel maintainers.

## 7.2. Trustworthy domain-specific compilers

To achieve safety and composability, we believe that an holistic approach is called for, involving not only the design of a domain-specific *syntax* but also of a domain-specific *semantics*. Concretely, we are exploring the design of *certified domain-specific compilers* that integrate, from the ground up, a denotational and domain-specific semantics as part of the design of a domain-specific language. This vision is illustrated by our work on the safe compilation of Coq programs into secure OCaml code [10]. It combines ideas from gradual typing – through which types are compiled into run-time assertions – and the theory of ornaments [31] – through which Coq datatypes can be related to OCaml datatypes. Within this formal framework, we enable a secure interaction, termed *dependent interoperability*, between correct-by-construction software and untrusted programs, be it system calls or legacy libraries. To do so, we trade static guarantees for runtime checks, thus allowing OCaml values to be safely coerced to dependently-typed Coq values and, conversely, to expose dependently-typed Coq programs defensively as OCaml programs. Our framework is developed in Coq: it is constructive and verified in the strictest sense of the terms. It thus becomes possible to internalize and hand-tune the extraction of dependently-typed programs to interoperable OCaml programs within Coq itself. This work is the result of a collaboration with Eric Tanter, from the University of Chile, and Nicolas Tabareau, from the Gallinette Inria project-team.

## 7.3. High-performance domain-specific compilers

As part of Darius Mercadier's PhD project, we are developing a synchronous dataflow language targeting high-performance (and, eventually, verified) implementations of bitsliced algorithms, with application to cryptographical algorithms [33]. Using our Usuba language, cryptographers can specify a block cipher at a very high level as a set of dataflow equations. From such a description, our `usubac` compiler is able to generate efficient, vectorized code exploiting the SIMD instruction sets of the underlying architecture. We have demonstrated that our generated code performs on par with hand-tuned assembly programs while, at the same time, being able to target multiple CPU architectures as well as multiple generations of SIMD instruction sets on each architecture. This project illustrates perfectly our methodology: the design of Usuba is driven by semantic considerations (bitslicing is only meaningful for bit parallel operations) that are then structured using types and subsequently reifed into syntactic artefacts. Our preliminary results [15], published in an international workshop, are encouraging.

## 7.4. Multicore schedulers

As a side-effect of our work on verification of schedulers [48], we have contributed to an analysis of the impact on application performance of the design and implementation choices made in two widely used open-source schedulers: ULE, the default FreeBSD scheduler, and CFS, the default Linux scheduler. In a paper published at USENIX ATC'18 [13], we compare ULE and CFS in otherwise identical circumstances. This work involves porting ULE to Linux, and using it to schedule all threads that are normally scheduled by CFS. We compare

the performance of a large suite of applications on the modified kernel running ULE and on the standard Linux kernel running CFS. The observed performance differences are solely the result of scheduling decisions, and do not reflect differences in other subsystems between FreeBSD and Linux. We found that there is no overall winner. On many workloads the two schedulers perform similarly, but for some workloads there are significant and even surprising differences. ULE may cause starvation, even when executing a single application with identical threads, but this starvation may actually lead to better application performance for some workloads. The more complex load balancing mechanism of CFS reacts more quickly to workload changes, but ULE achieves better load balance in the long run.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

- Orange Labs, 2016-2018, 120 000 euros. The purpose of this contract is to apply the techniques developed in the context of the PhD of Antoine Blin to the domain of Software Defined Networks where network functions are run using virtual machines on commodity multicore machines.
- Thales Research, 2016-2018, 45 000 euros. The purpose of this contract is to enable the usage of multicore architectures in avionics systems. The PhD of Cédric Courtaud is supported by a CIFRE fellowship as part of this contract.

## 8.2. Bilateral Grants with Industry

- Oracle, 2018-2019, 100 000 dollars. Operating system schedulers are often a performance bottleneck on multicore architectures because in order to scale, schedulers cannot make optimal decisions and instead have to rely on heuristics. Detecting that performance degradation comes from the scheduler level is extremely difficult because the issue has not been recognized until recently, and with traditional profilers, both the application and the scheduler affect the monitored metrics in the same way.

  The first objective of this project is to produce a profiler that makes it possible to find out whether a bottleneck during application runtime is caused by the application itself, by suboptimal OS scheduler behavior, or by a combination of the two. It will require understanding, analyzing and classifying performance bottlenecks that are caused by schedulers, and devising ways to detect them and to provide enough information for the user to understand the root cause of the issue. Following this, the second objective of this project is to use the profiler to better understand which kinds of workloads suffer from poor scheduling, and to propose new algorithms, heuristics and/or a new scheduler design that will improve the situation. Finally, the third contribution will be a methodology that makes it possible to track scheduling bottlenecks in a specific workload using the profiler, to understand them, and to fix them either at the application or at the scheduler level. We believe that the combination of these three contributions will make it possible to fully harness the power of multicore architectures for any workload.

# 9. Partnerships and Cooperations

## 9.1. Regional Initiatives

- City of Paris, 2016-2019, 100 000 euros. As part of the “Émergence - young team” program the city of Paris is supporting part of our work on domain-specific languages and trustworthy domain-specific compilers.

## 9.2. National Initiatives

### *9.2.1. ANR*

**ITrans** - awarded in 2016, duration 2017 - 2020

Members: LIP6 (Whisper), David Lo (Singapore Management University)

Coordinator: Julia Lawall

Whisper members: Julia Lawall, Gilles Muller, Lucas Serrano, Van-Anh Nguyen

Funding: ANR PRCI, 287,820 euros.

Objectives:

Large, real-world software must continually change, to keep up with evolving requirements, fix bugs, and improve performance, maintainability, and security. This rate of change can pose difficulties for clients, whose code cannot always evolve at the same rate. This project will target the problems of *forward porting*, where one software component has to catch up to a code base with which it needs to interact, and *back porting*, in which it is desired to use a more modern component in a context where it is necessary to continue to use a legacy code base, focusing on the context of Linux device drivers. In this project, we will take a *history-guided source-code transformation-based* approach, which automatically traverses the history of the changes made to a software system, to find where changes in the code to be ported are required, gathers examples of the required changes, and generates change rules to incrementally back port or forward port the code. Our approach will be a success if it is able to automatically back and forward port a large number of drivers for the Linux operating system to various earlier and later versions of the Linux kernel with high accuracy while requiring minimal developer effort. This objective is not achievable by existing techniques.

**VeriAmos** - awarded in 2018, duration 2018 - 2021

Members: Inria (Antique, Whisper), UGA (Erods)

Coordinator: Xavier Rival

Whisper members: Julia Lawall, Gilles Muller

Funding: ANR, 121,739 euros.

Objectives:

General-purpose Operating Systems, such as Linux, are increasingly used to support high-level functionalities in the safety-critical embedded systems industry with usage in automotive, medical and cyber-physical systems. However, it is well known that general purpose OSes suffer from bugs. In the embedded systems context, bugs may have critical consequences, even affecting human life. Recently, some major advances have been done in verifying OS kernels, mostly employing interactive theorem-proving techniques. These works rely on the formalization of the programming language semantics, and of the implementation of a software component, but require significant human intervention to supply the main proof arguments. The VeriAmos project will attack this problem by building on recent advances in the design of domain-specific languages and static analyzers for systems code. We will investigate whether the restricted expressiveness and the higher level of abstraction provided by the use of a DSL will make it possible to design static analyzers that can statically and fully automatically verify important classes of semantic properties on OS code, while retaining adequate performance of the OS service. As a specific use-case, the project will target I/O scheduling components.

## 9.3. International Initiatives

### *9.3.1. Inria International Labs*

- EPFL-Inria Lab Our work on scheduling [13] and on the Ipanema DSL [48] is done as part of the EPFL-Inria Lab. Our direct partners, Willy Zwaenepoel and Baptiste Lepers, have moved to the University of Sydney in September 2018. Therefore we have migrated our cooperation.

### 9.3.2. Inria International Partners

#### 9.3.2.1. Informal International Partners

- We collaborate with David Lo and Lingxiao Jiang of Singapore Management University, who are experts in software mining, clone detection, and information retrieval techniques. Our work with Lo and/or Jiang has led to 8 joint publications since 2013 [58], [69], [71], [74], [75], [76], [79], [77], at conferences including ASE and ICSME. The ITrans ANR is a joint project with them.
- We collaborate with David Lo and James Hoang of Singapore Management University and with Sasha Levin of Microsoft on the use of machine learning to identify stable-relevant patches in the Linux kernel. Preliminary results from this collaboration have been presented with Sasha Levin at the Open Source Summit North America, the Open Source Summit Europe, and the Linux Plumbers Conference kernel summit track.
- Our previous collaboration with EPFL has been transfered to the University of Sydney due to the moves of Willy Zwaenepoel and Baptiste Lepers.
- We collaborate with Christoph Reichenbach of the University of Lund and Krishna Narasimhan of Itemis (Germany) on program transformation and the design of tools for code clone management [11].
- We collaborate with Jia-Ju Bai of Tsinghua University on bug finding in Linux kernel code, particularly focusing on issues requiring interprocedural analysis [12].
- As part of the LIP6 Invited Professor program, we have initiated a collaboration between Karine Heydeman (ALSOC team – LIP6, France) and Patrick Schaumont (Virginia Tech, US) on the development of fault-resistant and side-channel attack resistant compilation techniques.

## 9.4. International Research Visitors

### 9.4.1. Visits of International Scientists

- Patrick Schaumont of Virginia Tech visited LIP6 in July and November 2018, as part of the LIP6 Invited Professor program.
- David Lo and Lingxiao Jiang of Singapore Management University visited the Whisper team for two weeks in October 2018 as part of the ANR ITrans project.
- Michele Martone of the Leibniz Supercomputing Centre in Munich Germany made two visits of one week each to the Whisper team in August and December to work on applying Coccinelle to high performance computing code.

#### 9.4.1.1. Internships

- Jonathan Carroll of Oberlin College spent January 2018 working on using machine learning to identify stable-relevant patches for the Linux kernel.
- David Bergvelt of the University of Illinois at Urbana-Champaign spent May-August 2018 working on applying Verifiable C, developed at Princeton, to verification of process schedulers.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### 10.1.1. Scientific Events Selection

#### 10.1.1.1. Chair of Conference Program Committees

- Gilles Muller: DSN 2018
- Julia Lawall: ASE 2018 Tool Demo track.

*10.1.1.2. Member of the Conference Program Committees*

- Gilles Muller: OSDI 2018, EuroSys 2018
- Julia Lawall: EuroSys 2018, ICSE-NIER 2018, ASPLOS 2018 ERC, PEPM 2018, SCAM 2018, APSys 2018, USENIX ATC 2018, CARI 2018

### *10.1.2. Journal*

*10.1.2.1. Member of the Editorial Boards*

- Julia Lawall: Editorial board of Science of Computer Programing (2008 - present).

*10.1.2.2. Reviewer - Reviewing Activities*

- Julia Lawall: Transactions on Software Engineering, Software: Evolution and Process, IEEE Transactions on Reliability, ACM Transactions on Embedded Computing Systems

### *10.1.3. Invited Talks*

- Gilles Muller:
  - "Provably Work Conserving Multicore Schedulers", University of Bordeaux, June 13, 2018.
  - "Safe multicore scheduling in a Linux cluster environment", 3rd GDR RSD and ASF Winter School on Distributed Systems and Networks, Sept Laux, March 20, 2018.
- Julia Lawall:
  - "Coccinelle: 10 Years of Automated Evolution in the Linux Kernel", 8th Inria/Technicolor Workshop On Systems, Rennes, December 11, 2018.
  - "Software evolution and bug finding using Coccinelle", Lightweight analysis and verification techniques, Verimag, Grenoble, December 11, 2018.
  - "Coccinelle: 10 Years of Automated Evolution in the Linux Kernel", Conférence d'informatique en Parallélisme, Architecture et Système (COMPAS), Toulouse, July 3, 2018.
  - "Coccinelle: Practical Program Transformation for the Linux Kernel", EJCP 2018 : École Jeunes Chercheurs et Jeunes Chercheuses en Programmation 2018, June 25, 2018.
  - "Introduction to Coccinelle and its usage in the Linux Kernel", Conférence MiNET, Telecom SudParis, May 24, 2018.
- Pierre-Évariste Dagand gave a seminar at the Collège de France entitled "Types dépendants : tout un programme" (November 28, 2018), as part of Xavier Leroy's chair "Sciences du logiciel".
- Lucas Serrano: "Inference of Semantic Patches from Code Examples", The Seventh International Workshop on Software Mining, with ASE, September 3, 2018.
- Cedric Courtaud "Toward an Efficient Data Plane for Memory Systems Interference Regulation in COTS Multi-core Systems", The NExt Step TOwards multi-core Real-time systems workshop, ULB, May 18, 2018.

### *10.1.4. Scientific Expertise*

- Julia Lawall was part of the midterm review panel of the NSF Expedition in Computing project DeepSpec.

### *10.1.5. Research Administration*

- Julia Lawall: IFIP TC secretary (2012 - present). Elected member of IFIP WG 2.11 (Program Generation).

  Member of a hiring committee for a Maître de conférences position at Université Paris Diderot

  Board member of Software Heritage (https://www.softwareheritage.org/).

- Gilles Muller: Elected member of IFIP WG 10.4 (Dependability), representative of Inria in Sorbonne University's advisory committee for research, member of the project committee board of the Inria Paris Center.
- Bertil Folliot: Elected member of the IFIP WG10.3 working group (Concurrent systems)

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

- Professional Licence: Bertil Folliot, Programmation C, L2, UPMC, France
- Professional Licence: Bertil Folliot, Lab projects, L2, UPMC, France
- Master: Pierre-Évariste Dagand, Specification and Validation of Programs, M2, UPMC, France
- Licence: Pierre-Évariste Dagand, INF311: Introduction to Programming, L1, École Polytechnique, France
- Master: Pierre-Évariste Dagand, INF559: Computer Architecture and Operating Systems, M1, École Polytechnique, France

### *10.2.2. Supervision*

- PhD : Mariem Saeid, soutenue le 25/9/2018, Jens Gustedt (Camus), Gilles Muller.
- PhD in progress : Cédric Courtaud, CIFRE Thalès, 2016-2019, Gilles Muller, Julien Sopéna (Delys).
- PhD in progress : Redha Gouicem, 2016-2019, Gilles Muller, Julien Sopéna (Delys).
- PhD in progress : Darius Mercadier, 2017-2020, Pierre-Évariste Dagand, Gilles Muller.
- PhD in progress : Lucas Serrano, 2017-2020, Julia Lawall.

### *10.2.3. Juries*

- Julia Lawall: PhD juries of Ferdian Thung, SMU (reporter), Thibaut Girka, Université Paris Diderot (president), Thomas Durieux, Lille (examiner).

## 10.3. Popularization

- Julia Lawall: Coordinator of the Outreachy internship program for the Linux kernel, until March 2018. Outreachy provides remote 3-month internships twice a year for women and other underrepresented minorities on open source projects. Julia Lawall also mentored Aishwarya Pant as part of this program.
- Julia Lawall, "Building Stable Trees with Machine Learning", Open Source Summit North America, August 2018, with Sasha Levin. Open Source Summit Europe, October 2018, with Sasha Levin. Linux Plumbers Conference, kernel summit track, November 2018, with Sasha Levin.
- Julia Lawall, "Coccinelle: 10 Years of Automated Evolution and Bug Finding in the Linux Kernel", Open Source Summit Europe, October 2018.
- Julia Lawall, "Panel Discussion: Outreachy Kernel Internship Report" (moderator), Open Source Summit Europe, October 2018.
- Julia Lawall, "Panel Discussion: An Exploration of Insights & Issues Related to Mentoring Programs" (participant), Open Source Summit Europe, October 2018.
- Julia Lawall, "Interprocedural Static Analysis Strategies for the Linux Kernel: Detecting SAC Bugs as an Example (Work in Progress)", Linux Kernel Real Time Summit, October 2018.
- Julia Lawall, "Kernel Panel" (participant), Linux Plumbers Conference, November 2018.

# 11. Bibliography

## Major publications by the team in recent years

[1] J. BRUNEL, D. DOLIGEZ, R. R. HANSEN, J. L. LAWALL, G. MULLER.*A foundation for flow-based program matching using temporal logic and model checking*, in "POPL", Savannah, GA, USA, ACM, January 2009, p. 114–126

[2] L. BURGY, L. RÉVEILLÈRE, J. L. LAWALL, G. MULLER.*Zebu: A Language-Based Approach for Network Protocol Message Processing*, in "IEEE Trans. Software Eng.", 2011, vol. 37, n$^{o}$ 4, p. 575-591

[3] P.-É. DAGAND, A. BAUMANN, T. ROSCOE.*Filet-o-Fish: practical and dependable domain-specific languages for OS development*, in "Programming Languages and Operating Systems (PLOS)", 2009, p. 51–55

[4] A. KENNEDY, N. BENTON, J. B. JENSEN, P.-É. DAGAND.*Coq: The World's Best Macro Assembler?*, in "PPDP", Madrid, Spain, ACM, 2013, p. 13–24

[5] G. MULLER, C. CONSEL, R. MARLET, L. P. BARRETO, F. MÉRILLON, L. RÉVEILLÈRE.*Towards Robust OSes for Appliances: A New Approach Based on Domain-specific Languages*, in "Proceedings of the 9th Workshop on ACM SIGOPS European Workshop: Beyond the PC: New Challenges for the Operating System", Kolding, Denmark, 2000, p. 19–24

[6] G. MULLER, J. L. LAWALL, H. DUCHESNE.*A Framework for Simplifying the Development of Kernel Schedulers: Design and Performance Evaluation*, in "HASE - High Assurance Systems Engineering Conference", Heidelberg, Germany, IEEE, October 2005, p. 56–65

[7] F. MÉRILLON, L. RÉVEILLÈRE, C. CONSEL, R. MARLET, G. MULLER.*Devil: An IDL for hardware programming*, in "Proceedings of the Fourth Symposium on Operating Systems Design and Implementation (OSDI)", San Diego, California, USENIX Association, October 2000, p. 17–30

[8] Y. PADIOLEAU, J. L. LAWALL, R. R. HANSEN, G. MULLER.*Documenting and Automating Collateral Evolutions in Linux Device Drivers*, in "EuroSys", Glasgow, Scotland, March 2008, p. 247–260

[9] N. PALIX, G. THOMAS, S. SAHA, C. CALVÈS, J. L. LAWALL, G. MULLER.*Faults in Linux 2.6*, in "ACM Transactions on Computer Systems", June 2014, vol. 32, n$^{o}$ 2, p. 4:1–4:40

## Publications of the year

### Articles in International Peer-Reviewed Journal

[10] P.-E. DAGAND, N. TABAREAU, É. TANTER.*Foundations of Dependent Interoperability*, in "Journal of Functional Programming", March 2018, vol. 28 [*DOI :* 10.1017/S0956796818000011], https://hal.inria.fr/hal-01629909

[11] K. NARASIMHAN, C. REICHENBACH, J. LAWALL.*Cleaning up Copy-Paste Clones with Interactive Merging*, in "Journal of Automated Software Engineering", August 2018, vol. 25, n$^{o}$ 3, p. 627-673 [*DOI :* 10.1007/s10515-018-0238-5], https://hal.inria.fr/hal-01853896

### International Conferences with Proceedings

[12] J.-J. BAI, Y.-P. WANG, J. LAWALL, S.-M. HU.*DSAC: Effective Static Analysis of Sleep-in-Atomic-Context Bugs in Kernel Modules*, in "2018 USENIX Annual Technical Conference", Boston, MA, United States, July 2018, https://hal.inria.fr/hal-01853268

[13] J. BOURON, S. CHEVALLEY, B. LEPERS, W. ZWAENEPOEL, R. GOUICEM, J. LAWALL, G. MULLER, J. SOPENA.*The Battle of the Schedulers: FreeBSD ULE vs. Linux CFS*, in "2018 USENIX Annual Technical Conference", Boston, MA, United States, July 2018, https://hal.inria.fr/hal-01853267

[14] J. LAWALL, G. MULLER.*Coccinelle: 10 Years of Automated Evolution in the Linux Kernel*, in "2018 USENIX Annual Technical Conference", Boston, MA, United States, July 2018, https://hal.inria.fr/hal-01853271

[15] D. MERCADIER, P.-É. DAGAND, L. LACASSAGNE, G. MULLER.*Usuba, Optimizing & Trustworthy Bitslicing Compiler*, in "WPMVP'18 - Workshop on Programming Models for SIMD/Vector Processing", Vienna, Austria, ACM Press, February 2018 [*DOI :* 10.1145/3178433.3178437], https://hal.archives-ouvertes.fr/hal-01657259

### Other Publications

[16] M. MARTONE, L. IAPICHINO, N. HAMMER, J. LAWALL.*Automating Data Layout Conversion in a Large Cosmological Simulation Code*, September 2018, CoSaS 2018 - International Symposium on Computational Science at Scale, Poster, https://hal.inria.fr/hal-01890314

## References in notes

[17] T. BALL, E. BOUNIMOVA, B. COOK, V. LEVIN, J. LICHTENBERG, C. MCGARVEY, B. ONDRUSEK, S. K. RAJAMANI, A. USTUNER.*Thorough Static Analysis of Device Drivers*, in "EuroSys", 2006, p. 73–85

[18] A. BAUMANN, P. BARHAM, P.-É. DAGAND, T. HARRIS, R. ISAACS, S. PETER, T. ROSCOE, A. SCHÜPBACH, A. SINGHANIA.*The multikernel: A new OS architecture for scalable multicore systems*, in "SOSP", 2009, p. 29–44

[19] T. F. BISSYANDÉ, L. RÉVEILLÈRE, J. L. LAWALL, Y.-D. BROMBERG, G. MULLER.*Implementing an embedded compiler using program transformation rules*, in "Software: Practice and Experience", 2013

[20] T. F. BISSYANDÉ, L. RÉVEILLÈRE, J. LAWALL, Y.-D. BROMBERG, G. MULLER.*Implementing an Embedded Compiler using Program Transformation Rules*, in "Software: Practice and Experience", February 2015, vol. 45, n° 2, p. 177-196, https://hal.archives-ouvertes.fr/hal-00844536

[21] T. F. BISSYANDÉ, L. RÉVEILLÈRE, J. LAWALL, G. MULLER.*Ahead of Time Static Analysis for Automatic Generation of Debugging Interfaces to the Linux Kernel*, in "Automated Software Engineering", May 2014, p. 1-39 [*DOI :* 10.1007/s10515-014-0152-4], https://hal.archives-ouvertes.fr/hal-00992283

[22] A. P. BLACK, S. DUCASSE, O. NIERSTRASZ, D. POLLET.*Pharo by Example*, Square Bracket Associates, 2010

[23] E. BRADY, K. HAMMOND.*Resource-Safe Systems Programming with Embedded Domain Specific Languages*, in "14th International Symposium on Practical Aspects of Declarative Languages (PADL)", LNCS, Springer, 2012, vol. 7149, p. 242–257

[24] T. BRAIBANT, D. POUS.*An Efficient Coq Tactic for Deciding Kleene Algebras*, in "1st International Conference on Interactive Theorem Proving (ITP)", LNCS, Springer, 2010, vol. 6172, p. 163–178

[25] C. CADAR, D. DUNBAR, D. R. ENGLER.*KLEE: Unassisted and Automatic Generation of High-Coverage Tests for Complex Systems Programs*, in "OSDI", 2008, p. 209–224

[26] V. CHIPOUNOV, G. CANDEA.*Reverse Engineering of Binary Device Drivers with RevNIC*, in "EuroSys", 2010, p. 167–180

[27] A. CHLIPALA.*The Bedrock Structured Programming System: Combining Generative Metaprogramming and Hoare Logic in an Extensible Program Verifier*, in "ICFP", 2013, p. 391–402

[28] L. A. CLARKE.*A system to generate test data and symbolically execute programs*, in "IEEE Transactions on Software Engineering", 1976, vol. 2, n$^{o}$ 3, p. 215–222

[29] E. CLARKE, O. GRUMBERG, S. JHA, Y. LU, H. VEITH.*Counterexample-guided abstraction refinement for symbolic model checking*, in "J. ACM", 2003, vol. 50, n$^{o}$ 5, p. 752–794

[30] P. COUSOT, R. COUSOT.*Abstract Interpretation: Past, Present and Future*, in "CSL-LICS", 2014, p. 2:1–2:10

[31] P.-É. DAGAND.*Reusability and Dependent Types*, University of Strathclyde, 2013

[32] I. DILLIG, T. DILLIG, A. AIKEN.*Sound, complete and scalable path-sensitive analysis*, in "PLDI", June 2008, p. 270–280

[33] D. DINU, Y. L. CORRE, D. KHOVRATOVICH, L. PERRIN, J. GROSSSCHÄDL, A. BIRYUKOV.*Triathlon of Lightweight Block Ciphers for the Internet of Things*, 2015, Cryptology ePrint Archive, Report 2015/209

[34] D. R. ENGLER, B. CHELF, A. CHOU, S. HALLEM.*Checking System Rules Using System-Specific, Programmer-Written Compiler Extensions*, in "OSDI", 2000, p. 1–16

[35] D. R. ENGLER, D. Y. CHEN, A. CHOU, B. CHELF.*Bugs as Deviant Behavior: A General Approach to Inferring Errors in Systems Code*, in "SOSP", 2001, p. 57–72

[36] A. GOLDBERG, D. ROBSON.*Smalltalk-80: The Language and Its Implementation*, Addison-Wesley, 1983

[37] L. GU, A. VAYNBERG, B. FORD, Z. SHAO, D. COSTANZO.*CertiKOS: A Certified Kernel for Secure Cloud Computing*, in "Proceedings of the Second Asia-Pacific Workshop on Systems (APSys)", 2011, p. 3:1–3:5

[38] L. GUO, J. L. LAWALL, G. MULLER.*Oops! Where did that code snippet come from?*, in "11th Working Conference on Mining Software Repositories, MSR", Hyderabad, India, ACM, May 2014, p. 52–61

[39] A. ISRAELI, D. G. FEITELSON.*The Linux kernel as a case study in software evolution*, in "Journal of Systems and Software", 2010, vol. 83, n$^{o}$ 3, p. 485–501

[40] A. KADAV, M. M. SWIFT.*Understanding modern device drivers*, in "ASPLOS", 2012, p. 87–98

[41] G. A. KILDALL.*A Unified Approach to Global Program Optimization*, in "POPL", 1973, p. 194–206

[42] G. KLEIN, K. ELPHINSTONE, G. HEISER, J. ANDRONICK, D. COCK, P. DERRIN, D. ELKADUWE, K. ENGELHARDT, R. KOLANSKI, M. NORRISH, T. SEWELL, H. TUCH, S. WINWOOD.*seL4: formal verification of an OS kernel*, in "SOSP", 2009, p. 207–220

[43] J. L. LAWALL, J. BRUNEL, N. PALIX, R. R. HANSEN, H. STUART, G. MULLER.*WYSIWIB: Exploiting fine-grained program structure in a scriptable API-usage protocol-finding process*, in "Software, Practice Experience", 2013, vol. 43, n$^{o}$ 1, p. 67–92

[44] J. L. LAWALL, B. LAURIE, R. R. HANSEN, N. PALIX, G. MULLER.*Finding Error Handling Bugs in OpenSSL using Coccinelle*, in "Proceeding of the 8th European Dependable Computing Conference (EDCC)", Valencia, Spain, April 2010, p. 191–196

[45] J. L. LAWALL, D. LO.*An automated approach for finding variable-constant pairing bugs*, in "25th IEEE/ACM International Conference on Automated Software Engineering", Antwerp, Belgium, September 2010, p. 103–112

[46] J. LAWALL, D. PALINSKI, L. GNIRKE, G. MULLER.*Fast and Precise Retrieval of Forward and Back Porting Information for Linux Device Drivers*, in "2017 USENIX Annual Technical Conference", Santa Clara, CA, United States, July 2017, 12, https://hal.inria.fr/hal-01556589

[47] C. LE GOUES, W. WEIMER.*Specification Mining with Few False Positives*, in "TACAS", York, UK, Lecture Notes in Computer Science, March 2009, vol. 5505, p. 292–306

[48] B. LEPERS, W. ZWAENEPOEL, J.-P. LOZI, N. PALIX, R. GOUICEM, J. SOPENA, J. LAWALL, G. MULLER.*Towards Proving Optimistic Multicore Schedulers*, in "HotOS 2017 - 16th Workshop on Hot Topics in Operating Systems", Whistler, British Columbia, Canada, ACM SIGOPS, May 2017, 6 [*DOI :* 10.1145/3102980.3102984], https://hal.inria.fr/hal-01556597

[49] Z. LI, S. LU, S. MYAGMAR, Y. ZHOU.*CP-Miner: A Tool for Finding Copy-paste and Related Bugs in Operating System Code*, in "OSDI", 2004, p. 289–302

[50] Z. LI, Y. ZHOU.*PR-Miner: automatically extracting implicit programming rules and detecting violations in large software code*, in "Proceedings of the 10th European Software Engineering Conference", 2005, p. 306–315

[51] D. LO, S. KHOO.*SMArTIC: towards building an accurate, robust and scalable specification miner*, in "FSE", 2006, p. 265–275

[52] J.-P. LOZI, F. DAVID, G. THOMAS, J. LAWALL, G. MULLER.*Fast and Portable Locking for Multicore Architectures*, in "ACM Transactions on Computer Systems", January 2016 [*DOI :* 10.1145/2845079], https://hal.inria.fr/hal-01252167

[53] S. LU, S. PARK, Y. ZHOU.*Finding Atomicity-Violation Bugs through Unserializable Interleaving Testing*, in "IEEE Transactions on Software Engineering", 2012, vol. 38, n$^{o}$ 4, p. 844–860

[54] M. MERNIK, J. HEERING, A. M. SLOANE.*When and How to Develop Domain-specific Languages*, in "ACM Comput. Surv.", December 2005, vol. 37, n$^{o}$ 4, p. 316–344, http://dx.doi.org/10.1145/1118890.1118892

[55] G. MORRISETT, G. TAN, J. TASSAROTTI, J.-B. TRISTAN, E. GAN.*RockSalt: better, faster, stronger SFI for the x86*, in "PLDI", 2012, p. 395-404

[56] M. ODERSKY, T. ROMPF.*Unifying functional and object-oriented programming with Scala*, in "Commun. ACM", 2014, vol. 57, n$^{o}$ 4, p. 76–86

[57] M. C. OLESEN, R. R. HANSEN, J. L. LAWALL, N. PALIX.*Coccinelle: Tool support for automated CERT C Secure Coding Standard certification*, in "Science of Computer Programming", October 2014, vol. 91, n$^{o}$ B, p. 141–160, https://hal.inria.fr/hal-01096185

[58] K. PAVNEET SINGH, F. THUNG, D. LO, J. LAWALL.*An Empirical Study on the Adequacy of Testing in Open Source Projects*, in "21st Asia-Pacific Software Engineering Conference", Jeju, South Korea, December 2014, https://hal.inria.fr/hal-01096132

[59] T. REPS, T. BALL, M. DAS, J. LARUS.*The Use of Program Profiling for Software Maintenance with Applications to the Year 2000 Problem*, in "ESEC/FSE", 1997, p. 432–449

[60] L. R. RODRIGUEZ, J. LAWALL.*Increasing Automation in the Backporting of Linux Drivers Using Coccinelle*, in "11th European Dependable Computing Conference - Dependability in Practice", Paris, France, 11th European Dependable Computing Conference - Dependability in Practice, November 2015, https://hal.inria.fr/hal-01213912

[61] C. RUBIO-GONZÁLEZ, H. S. GUNAWI, B. LIBLIT, R. H. ARPACI-DUSSEAU, A. C. ARPACI-DUSSEAU.*Error propagation analysis for file systems*, in "PLDI", Dublin, Ireland, ACM, June 2009, p. 270–280

[62] L. RYZHYK, P. CHUBB, I. KUZ, E. LE SUEUR, G. HEISER.*Automatic device driver synthesis with Termite*, in "SOSP", 2009, p. 73–86

[63] L. RYZHYK, A. WALKER, J. KEYS, A. LEGG, A. RAGHUNATH, M. STUMM, M. VIJ.*User-Guided Device Driver Synthesis*, in "OSDI", 2014, p. 661–676

[64] R. K. SAHA, J. L. LAWALL, S. KHURSHID, D. E. PERRY.*On the Effectiveness of Information Retrieval Based Bug Localization for C Programs*, in "ICSME 2014 - 30th International Conference on Software Maintenance and Evolution", Victoria, Canada, IEEE, September 2014, p. 161-170 [*DOI :* 10.1109/ICSME.2014.38], https://hal.inria.fr/hal-01086082

[65] R. SAHA, J. L. LAWALL, S. KHURSHID, D. E. PERRY.*On the Effectiveness of Information Retrieval based Bug Localization for C Programs*, in "International Conference on Software Maintenance and Evolution (ICSME)", Victoria, BC, Canada, September 2014

[66] S. SAHA, J.-P. LOZI, G. THOMAS, J. LAWALL, G. MULLER.*Hector: Detecting resource-release omission faults in error-handling code for systems software*, in "DSN 2013 - 43rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)", Budapest, Hungary, IEEE Computer Society, June 2013, p. 1-12 [*DOI :* 10.1109/DSN.2013.6575307], https://hal.inria.fr/hal-00918079

[67] D. A. SCHMIDT.*Data Flow Analysis is Model Checking of Abstract Interpretations*, in "POPL", 1998, p. 38–48

[68] P. SENNA TSCHUDIN, J. LAWALL, G. MULLER.*3L: Learning Linux Logging*, in "BElgian-NEtherlands software eVOLution seminar (BENEVOL 2015)", Lille, France, December 2015, https://hal.inria.fr/hal-01239980

[69] P. SENNA TSCHUDIN, L. RÉVEILLÈRE, L. JIANG, D. LO, J. LAWALL, G. MULLER.*Understanding the genetic makeup of Linux device drivers*, in "PLOS'13 - 7th Workshop on Programming Languages and

Operating Systems", Nemacolin Woodlands Resort, Pennsylvania, United States, ACM, November 2013 [*DOI :* 10.1145/2525528.2525536], https://hal.inria.fr/hal-00927070

[70] M. SHAPIRO.*Purpose-built languages*, in "Commun. ACM", 2009, vol. 52, n$^{o}$ 4, p. 36–41

[71] P. SINGH KOCHHAR, D. LO, J. LAWALL, N. NAGAPPAN.*Code Coverage and Postrelease Defects: A Large-Scale Study on Open Source Projects*, in "IEEE Transactions on Reliability", December 2017, vol. 66, n$^{o}$ 4, p. 1213 - 1228 [*DOI :* 10.1109/TR.2017.2727062], https://hal.inria.fr/hal-01653728

[72] R. TARTLER, D. LOHMANN, J. SINCERO, W. SCHRÖDER-PREIKSCHAT.*Feature consistency in compile-time-configurable system software: facing the Linux 10,000 feature problem*, in "EuroSys", 2011, p. 47–60

[73] F. THUNG, D. X. B. LE, D. LO, J. LAWALL.*Recommending Code Changes for Automatic Backporting of Linux Device Drivers*, in "32nd IEEE International Conference on Software Maintenance and Evolution (ICSME)", Raleigh, North Carolina, United States, IEEE, October 2016, https://hal.inria.fr/hal-01355859

[74] F. THUNG, D. LO, J. L. LAWALL.*Automated library recommendation*, in "WCRE 2013 - 20th Working Conference on Reverse Engineering", Koblenz, Germany, R. LÄMMEL, R. OLIVETO, R. ROBBES (editors), IEEE, October 2013, p. 182-191 [*DOI :* 10.1109/WCRE.2013.6671293], https://hal.inria.fr/hal-00918076

[75] F. THUNG, S. WANG, D. LO, J. LAWALL.*Automatic recommendation of API methods from feature requests*, in "ASE 2013 - 28th IEEE/ACM International Conference on Automated Software Engineering", Palo Alto, California, United States, E. DENNEY, T. BULTAN, A. ZELLER (editors), IEEE, November 2013, https://hal.inria.fr/hal-00918828

[76] Y. TIAN, D. LO, J. LAWALL.*Automated construction of a software-specific word similarity database*, in "2014 Software Evolution Week - IEEE Conference on Software Maintenance, Reengineering, and Reverse Engineering, CSMR-WCRE", Antwerp, Belgium, IEEE, February 2014, p. 44-53, https://hal.inria.fr/hal-01086077

[77] Y. TIAN, D. LO, J. LAWALL.*SEWordSim: software-specific word similarity database*, ACM, May 2014, p. 568-571, ICSE Companion 2014 - Companion Proceedings of the 36th International Conference on Software Engineering, Poster [*DOI :* 10.1145/2591062.2591071], https://hal.inria.fr/hal-01086079

[78] W. WANG, M. GODFREY.*A Study of Cloning in the Linux SCSI Drivers*, in "Source Code Analysis and Manipulation (SCAM)", IEEE, 2011

[79] S. WANG, D. LO, J. LAWALL.*Compositional Vector Space Models for Improved Bug Localization*, in "30th International Conference on Software Maintenance and Evolution", Victoria, Canada, IEEE, September 2014, p. 171-180, https://hal.inria.fr/hal-01086084

[80] J. YANG, C. HAWBLITZEL.*Safe to the Last Instruction: Automated Verification of a Type-safe Operating System*, in "PLDI", 2010, p. 99–110

# Project-Team WILLOW

## Models of visual object recognition and scene understanding

IN COLLABORATION WITH: Département d'Informatique de l'Ecole Normale Supérieure

IN PARTNERSHIP WITH:
**CNRS**

**Ecole normale supérieure de Paris**

RESEARCH CENTER
**Paris**

THEME
**Vision, perception and multimedia interpretation**

# Table of contents

# Project-Team WILLOW

*Creation of the Project-Team: 2007 June 01*

**Keywords:**

**Computer Science and Digital Science:**

A3.1.1. - Modeling, representation
A3.4. - Machine learning and statistics
A5.3. - Image processing and analysis
A5.4. - Computer vision
A9. - Artificial intelligence
A9.1. - Knowledge
A9.2. - Machine learning

**Other Research Topics and Application Domains:**

B9.5.1. - Computer science
B9.5.6. - Data science

# 1. Team, Visitors, External Collaborators

**Research Scientists**

Jean Ponce [Team leader, Inria, Senior Researcher, on leave from Ecole Normale Supérieure]
Ivan Laptev [Inria, Senior Researcher, HDR]
Josef Sivic [Inria, Senior Researcher, HDR]

**External Collaborator**

Mathieu Aubry [Ecole Nationale des Ponts et Chaussees]

**Technical Staff**

Sofiane Allayen [Inria, from May 2018]
Mauricio Diaz Melo [Inria, until Mar 2018]
Igor Kalevatykh [Inria]

**PhD Students**

Jean-Baptiste Alayrac [Inria, until Aug 2018]
Guilhem Cheron [Inria, until Aug 2018]
Theophile Dalens [Inria]
Thomas Eboli [Ecole Normale Superieure Paris]
Yana Hasson [Inria]
Yann Labbe [Ecole Normale Superieure Cachan, Intern from Apr 2018 then Phd from Sep 2018]
Zongmian Li [Inria]
Antoine Miech [Inria]
Julia Peyre [Inria]
Ronan Riochet [Inria]
Ignacio Rocco Spremolla [Inria]
Robin Strudel [Ecole Normale Superieure Paris, Intern from Apr 2018 then Phd from Oct 2018]
Matthew Trager [Inria, until Jun 2018]
Gul Varol [Inria]
Tuan Hung Vu [Inria, until Mar 2018]
Van Huy Vo [Ecole Normale Superieure Paris, from Dec 2018]
Dimitri Zhukov [Inria]

**Post-Doctoral Fellows**

Justin Carpentier [Inria, from Sep 2018]
Vijay Kumar Reddy [Inria, from Jul 2018]
Sergey Zagoruyko [Inria, from Mar 2018]

**Visiting Scientists**

Alexei Efros [UC Berkeley, from May 2018 until Jun 2018]
Ramazan Cinbis [Middle East Technical University, from Jul 2018 until Aug 2018]
David Fouhey [University of Michigan, from Sep 2018 until Nov 2018]
Pierre-Yves Masse [Czech Technical University, from Apr 2018]
Akihiko Torii [Tokyo Institute of Technology, from Apr 2018]

**Administrative Assistants**

Helene Bessin Rousseau [Inria, from Mar 2018]
Sabrine Boumizy [Inria, until Feb 2018]
Helene Milome [Inria, from Dec 2018]

# 2. Overall Objectives

## 2.1. Statement

Object recognition —or, in a broader sense, scene understanding— is the ultimate scientific challenge of computer vision: After 40 years of research, robustly identifying the familiar objects (chair, person, pet), scene categories (beach, forest, office), and activity patterns (conversation, dance, picnic) depicted in family pictures, news segments, or feature films is still beyond the capabilities of today's vision systems. On the other hand, truly successful object recognition and scene understanding technology will have a broad impact in application domains as varied as defense, entertainment, health care, human-computer interaction, image retrieval and data mining, industrial and personal robotics, manufacturing, scientific image analysis, surveillance and security, and transportation.

Despite the limitations of today's scene understanding technology, tremendous progress has been accomplished in the past ten years, due in part to the formulation of object recognition as a statistical pattern matching problem. The emphasis is in general on the features defining the patterns and on the algorithms used to learn and recognize them, rather than on the representation of object, scene, and activity categories, or the integrated interpretation of the various scene elements. WILLOW complements this approach with an ambitious research program explicitly addressing the representational issues involved in object recognition and, more generally, scene understanding.

Concretely, our objective is to develop geometric, physical, and statistical models for all components of the image interpretation process, including illumination, materials, objects, scenes, and human activities. These models will be used to tackle fundamental scientific challenges such as three-dimensional (3D) object and scene modeling, analysis, and retrieval; human activity capture and classification; and category-level object and scene recognition. They will also support applications with high scientific, societal, and/or economic impact in domains such as quantitative image analysis in science and humanities; film post-production and special effects; and video annotation, interpretation, and retrieval. Machine learning is a key part of our effort, with a balance of practical work in support of computer vision application and methodological research aimed at developing effective algorithms and architectures.

WILLOW was created in 2007: It was recognized as an Inria team in January 2007, and as an official project-team in June 2007. WILLOW is a joint research team between Inria Paris, Ecole Normale Supérieure (ENS) and Centre National de la Recherche Scientifique (CNRS).

This year we have hired three new Phd students: Yann Labbe (ENS Cachan), Robin Strudel (ENS Lyon) and Van Huy Vo. Alexei Efros (Professor, UC Berkeley, USA) visited Willow during May-June. Ramazan Cinbis (Middle East Technical University) and David Fouhey (University of Michigan) visited Willow in July-August and September-November, respectively. Akihiko Torii (Tokyo Institute of Technology) spent sabbatical at Willow from Apr to August 2018. Finally, Pierre-Yves Masse (post-doc, Czech Technical University) spent 50% of his time at Sierra (F. Bach) and Willow teams as a visiting post-doc within the framework of collaboration with the Intelligent Machine Perception project lead by J. Sivic at the Czech Technical University in Prague.

# 3. Research Program

## 3.1. 3D object and scene modeling, analysis, and retrieval

This part of our research focuses on geometric models of specific 3D objects at the local (differential) and global levels, physical and statistical models of materials and illumination patterns, and modeling and retrieval of objects and scenes in large image collections. Our past work in these areas includes research aimed at recognizing rigid 3D objects in cluttered photographs taken from arbitrary viewpoints (Rothganger *et al.*, 2006), segmenting video sequences into parts corresponding to rigid scene components before recognizing these in new video clips (Rothganger *et al.*, 2007), retrieval of particular objects and buildings from images and videos (Sivic and Zisserman, 2003) and (Philbin *et al.*, 2007), and a theoretical study of a general formalism for modeling central and non-central cameras using the formalism and terminology of classical projective geometry (Ponce, 2009 and Batog *et al.*, 2010).

We have also developed multi-view stereopsis algorithms that have proven remarkably effective at recovering intricate details and thin features of compact objects and capturing the overall structure of large-scale, cluttered scenes. We have obtained a US patent 8,331,615 [0] for the corresponding software (PMVS, https://github.com/pmoulon/CMVS-PMVS) which is available under a GPL license and used for film production by ILM and Weta as well as by Google in Google Maps. It is also the basic technology used by Iconem, a start-up founded by Y. Ubelmann, a Willow collaborator. We have also applied our multi-view-stereo approach to model archaeological sites together with developing representations and efficient retrieval techniques to enable matching historical paintings to 3D models of archaeological sites (Russel *et al.*, 2011).

Our current efforts in this area are outlined in detail in Section. 7.1.

## 3.2. Category-level object and scene recognition

The objective in this core part of our research is to learn and recognize quickly and accurately thousands of visual categories, including materials, objects, scenes, and broad classes of temporal events, such as patterns of human activities in picnics, conversations, etc. The current paradigm in the vision community is to model/learn one object category (read 2D aspect) at a time. If we are to achieve our goal, we have to break away from this paradigm, and develop models that account for the tremendous variability in object and scene appearance due to texture, material, viewpoint, and illumination changes within each object category, as well as the complex and evolving relationships between scene elements during the course of normal human activities.

Our current work in this area is outlined in detail in Section 7.2.

[0]The patent: "Match, Expand, and Filter Technique for Multi-View Stereopsis" was issued December 11, 2012 and assigned patent number 8,331,615.

## 3.3. Image restoration, manipulation and enhancement

The goal of this part of our research is to develop models, and methods for image/video restoration, manipulation and enhancement. The ability to “intelligently" manipulate the content of images and video is just as essential as high-level content interpretation in many applications: This ranges from restoring old films or removing unwanted wires and rigs from new ones in post production, to cleaning up a shot of your daughter at her birthday party, which is lovely but noisy and blurry because the lights were out when she blew the candles, or editing out a tourist from your Roman holiday video. Going beyond the modest abilities of current “digital zoom" (bicubic interpolation in general) so you can close in on that birthday cake, “deblock" a football game on TV, or turn your favorite DVD into a blue-ray, is just as important.

In this context, we believe there is a new convergence between computer vision, machine learning, and signal processing. For example: The idea of exploiting self-similarities in image analysis, originally introduced in computer vision for texture synthesis applications (Efros and Leung, 1999), is the basis for non-local means (Buades *et al.*, 2005), one of today’s most successful approaches to image restoration. In turn, by combining a powerful sparse coding approach to non-local means (Dabov *et al.*, 2007) with modern machine learning techniques for dictionary learning (Mairal *et al.*, 2010), we have obtained denoising and demosaicking results that are the state of the art on standard benchmarks (Mairal *et al.*, 2009).

Our current work is outlined in detail in Section 7.3.

## 3.4. Human activity capture and classification

From a scientific point of view, visual action understanding is a computer vision problem that until recently has received little attention outside of extremely specific contexts such as surveillance or sports. Many of the current approaches to the visual interpretation of human activities are designed for a limited range of operating conditions, such as static cameras, fixed scenes, or restricted actions. The objective of this part of our project is to attack the much more challenging problem of understanding actions and interactions in unconstrained video depicting everyday human activities such as in sitcoms, feature films, or news segments. The recent emergence of automated annotation tools for this type of video data (Everingham, Sivic, Zisserman, 2006; Laptev, Marszałek, Schmid, Rozenfeld, 2008; Duchenne, Laptev, Sivic, Bach, Ponce, 2009) means that massive amounts of labelled data for training and recognizing action models will at long last be available.

Our research agenda in this scientific domain is described below and our recent results are outlined in detail in Section 7.4.

- **Weakly-supervised learning and annotation of human actions in video.** We aim to leverage the huge amount of video data using readily-available annotations in the form of video scripts. Scripts, however, often provide only imprecise and incomplete information about the video. We address this problem with weakly-supervised learning techniques both at the text and image levels.
- **Descriptors for video representation.** Video representation has a crucial role for recognizing human actions and other components of a visual scene. Our work in this domain aims to develop generic methods for representing video data based on realistic assumptions. In particular, we develop deep learning methods and design new trainable representations for various tasks such as human action recognition, person detection, segmentation and tracking.

# 4. Application Domains

## 4.1. Introduction

We believe that foundational modeling work should be grounded in applications. This includes (but is not restricted to) the following high-impact domains.

## 4.2. Quantitative image analysis in science and humanities

We plan to apply our 3D object and scene modeling and analysis technology to image-based modeling of human skeletons and artifacts in anthropology, and large-scale site indexing, modeling, and retrieval in archaeology and cultural heritage preservation. Most existing work in this domain concentrates on image-based rendering, that is, the synthesis of good-looking pictures of artifacts and digs. We plan to focus instead on quantitative applications. We are engaged in a project involving the archaeology laboratory at ENS and focusing on image-based artifact modeling and decorative pattern retrieval in Pompeii. Application of our 3D reconstruction technology is now being explored in the field of cultural heritage and archeology by the start-up Iconem, founded by Y. Ubelmann, a Willow collaborator.

## 4.3. Video Annotation, Interpretation, and Retrieval

Both specific and category-level object and scene recognition can be used to annotate, augment, index, and retrieve video segments in the audiovisual domain. The Video Google system developed by Sivic and Zisserman (2005) for retrieving shots containing specific objects is an early success in that area. A sample application, suggested by discussions with Institut National de l'Audiovisuel (INA) staff, is to match set photographs with actual shots in film and video archives, despite the fact that detailed timetables and/or annotations are typically not available for either medium. Automatically annotating the shots is of course also relevant for archives that may record hundreds of thousands of hours of video. Some of these applications will be pursued in our MSR-Inria project.

# 5. Highlights of the Year

## 5.1. Highlights of the Year

### *5.1.1. Prizes and Awards*

Antoine Miech, winner of a 2018 Google Fellowship.

### *5.1.2. Visibility*

- J. Ponce co-organized the PRAIRIE AI Summer School, Grenoble, 2018, which brought together 200 participants representing 44 different nationalities, and selected from 700 applications, with 60% students, 15% academics , and 25% industrials. 25% of these participants were women.
- I. Laptev served as Program Chair for the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018. CVPR is the largest computer vision conference. The 2018 edition has 3,309 paper submissions, 979 accepted papers and 6,128 registered attendees.
- J. Ponce has been a key person in creating the PRAIRIE Institute for AI research in Paris, announced on the occasion of the AI for Humanity summit organized by President Emmanuel Macron in 2018 (https://www.inria.fr/en/news/news-from-inria/launch-of-the-prairie-institute). He has also been a key player in bringing together its industrial and international partners.

# 6. New Software and Platforms

## 6.1. NCNet

*Neighbourhood Consensus Networks*
KEYWORDS: Computer vision - Machine learning

FUNCTIONAL DESCRIPTION: Open source release of the software package for the NIPS'18 paper by Rocco et al. "Neighbourhood Consensus Networks". This release provides a full implementation of the method, including code for training models, and testing on standard datasets, as well as trained models.

- Participants: Ignacio Rocco Spremolla, Mircea Cimpoi, Akihiko Torii, Relja Arandjelovic, Tomas Pajdla and Josef Sivic
- Contact: Ignacio Rocco Spremolla
- Publication: Neighbourhood Consensus Networks
- URL: https://www.di.ens.fr/willow/research/ncnet/

## 6.2. Mixture-of-Embedding-Experts

KEYWORD: Computer vision
FUNCTIONAL DESCRIPTION: Joint understanding of video and language is an active research area with many applications. Prior work in this domain typically relies on learning text-video embeddings. One difficulty with this approach, however, is the lack of large-scale annotated video-caption datasets for training. To address this issue, we aim at learning text-video embeddings from heterogeneous data sources. To this end, we propose a Mixture-of-Embedding-Experts (MEE) model with ability to handle missing input modalities during training. As a result, our framework can learn improved text-video embeddings simultaneously from image and video datasets. We also show the generalization of MEE to other input modalities such as face descriptors.

- Participants: Ivan Laptev and Josef Sivic
- Contact: Antoine Miech
- Publication: Learning a Text-Video Embedding from Incomplete and Heterogeneous Data
- URL: https://www.di.ens.fr/willow/research/mee/

## 6.3. BodyNet

*BodyNet: Volumetric Inference of 3D Human Body Shapes*
KEYWORDS: Computer vision - Machine learning
FUNCTIONAL DESCRIPTION: BodyNet has the code to train multi-task neural networks for 2D/3D pose estimation, 2D body part segmentation, and 3D volumetric shape estimation of human bodies given single RGB images as input. The release also contains pre-trained models.

- Participants: Gül Varol Simsekli, Ivan Laptev and Cordelia Schmid
- Contact: Gül Varol Simsekli
- Publication: BodyNet: Volumetric Inference of 3D Human Body Shapes
- URL: https://www.di.ens.fr/willow/research/bodynet/

## 6.4. FlexWLoc

*Flexible Weakly supervised action Localization model*
KEYWORDS: Computer vision - Machine learning
FUNCTIONAL DESCRIPTION: Open source release of the software package for the NIPS'18 paper by Chéron et al. "A flexible model for training action localization with varying levels of supervision". This release provides a full implementation of the method, including code for training and testing.

- Participants: Guilhem Chéron, Jean-Baptiste Alayrac, Ivan Laptev and Cordelia Schmid
- Contact: Guilhem Chéron
- Publication: A flexible model for training action localization with varying levels of supervision
- URL: https://www.di.ens.fr/willow/research/weakactionloc/

## 6.5. Pinocchio

KEYWORDS: Robotics - Biomechanics - Mechanical multi-body systems
FUNCTIONAL DESCRIPTION: Pinocchio instantiates state-of-the-art Rigid Body Algorithms for poly-articulated systems based on revisited Roy Featherstone's algorithms. In addition, Pinocchio instantiates analytical derivatives of the main Rigid-Body Algorithms like the Recursive Newton-Euler Algorithms or the Articulated-Body Algorithm. Pinocchio is first tailored for legged robotics applications, but it can be used in extra contexts. It is built upon Eigen for linear algebra and FCL for collision detection. Pinocchio comes with a Python interface for fast code prototyping.

- Partner: CNRS
- Contact: JUSTIN CARPENTIER
- URL: https://github.com/stack-of-tasks/pinocchio

## 6.6. weakalign

*End-to-end weakly-supervised semantic alignment*
KEYWORDS: Computer vision - Machine learning
FUNCTIONAL DESCRIPTION: Open source release of the software package for the CVPR'18 paper by Rocco et al. "End-to-end weakly-supervised semantic alignment". This release provides a full implementation of the method, including code for training models, and testing on standard datasets, as well as trained models.

- Participants: Ignacio Rocco Spremolla, Relja Arandjelovic and Josef Sivic
- Contact: Ignacio Rocco Spremolla
- Publication: End-to-end weakly-supervised semantic alignment
- URL: https://www.di.ens.fr/willow/research/weakalign/

## 6.7. InLoc

*Indoor Visual Localization with Dense Matching and View Synthesis*
KEYWORD: Computer vision
FUNCTIONAL DESCRIPTION: Open source release of the software package for the CVPR'18 paper by Taira et al. "Indoor Visual Localization with Dense Matching and View Synthesis". This release provides a full implementation of the method.

- Participants: Hajime Taira, Masatoshi Okutomi, Torsten Sattler, Mircea Cimpoi, Marc Pollefeys, Josef Sivic, Tomas Pajdla and Akihiko Torii
- Contact: Josef Sivic
- Publication: InLoc: Indoor Visual Localization with Dense Matching and View Synthesis
- URL: https://github.com/HajimeTaira/InLoc_demo

# 7. New Results

## 7.1. 3D object and scene modeling, analysis, and retrieval

### *7.1.1. Indoor Visual Localization with Dense Matching and View Synthesis*

**Participants:** Hajime Taira, Masatoshi Okutomi, Torsten Sattler, Mircea Cimpoi, Marc Pollefeys, Josef Sivic, Tomas Pajdla, Akihiko Torii.

In [20], we seek to predict the 6 degree-of-freedom (6DoF) pose of a query photograph with respect to a large indoor 3D map. The contributions of this work are three-fold. First, we develop a new large-scale visual localization method targeted for indoor environments. The method proceeds along three steps: (i) efficient retrieval of candidate poses that ensures scalability to large-scale environments, (ii) pose estimation using dense matching rather than local features to deal with textureless indoor scenes, and (iii) pose verification by virtual view synthesis to cope with significant changes in viewpoint, scene layout, and occluders. Second, we collect a new dataset with reference 6DoF poses for large-scale indoor localization. Query photographs are captured by mobile phones at a different time than the reference 3D map, thus presenting a realistic indoor localization scenario. Third, we demonstrate that our method significantly outperforms current state-of-the-art indoor localization approaches on this new challenging data. Figure 1 presents some example results.

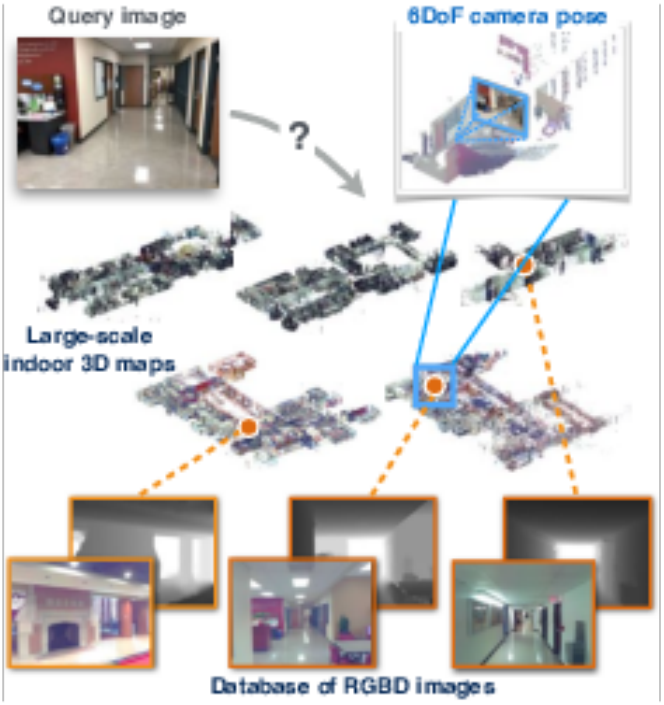


*Figure 1. Large-scale indoor visual localization. Given a database of geometrically-registered RGBD images, we predict the 6DoF camera pose of a query RGB image by retrieving candidate images, estimating candicate camera poses, and selecting the best matching camera pose. To address inherent difficulties in indoor visual localization, we introduce the ?InLoc? approach that performs a sequence of progressively stricter verification steps.*

### *7.1.2. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions*

**Participants:** Torsten Sattler, Will Maddern, Carl Toft, Akihiko Torii, Lars Hammarstrand, Erik Stenborg, Daniel Safari, Masatoshi Okutomi, Marc Pollefeys, Josef Sivic, Frederik Kahl, Tomas Pajdla.

Visual localization enables autonomous vehicles to navigate in their surroundings and augmented reality applications to link virtual to real worlds. Practical visual localization approaches need to be robust to a wide variety of viewing condition, including day-night changes, as well as weather and seasonal variations, while providing highly accurate 6 degree-of-freedom (6DOF) camera pose estimates. In [19], we introduce the first benchmark datasets specifically designed for analyzing the impact of such factors on visual localization. Using carefully created ground truth poses for query images taken under a wide variety of conditions, we evaluate the impact of various factors on 6DOF camera pose estimation accuracy through extensive experiments with state-of-the-art localization approaches. Based on our results, we draw conclusions about the difficulty of different conditions, showing that long-term localization is far from solved, and propose promising avenues for future work, including sequence-based localization approaches and the need for better local features. Our benchmark is available at visuallocalization.net. Figure 2 presents some example results.

*Figure 2. Visual localization in changing urban conditions. We present three new datasets, Aachen Day-Night, RobotCar Seasons (shown) and CMU Seasons for evaluating 6DOF localization against a prior 3D map (top) using registered query images taken from a wide variety of conditions (bottom), including day-night variation, weather, and seasonal changes over long periods of time.*

### *7.1.3. Changing Views on Curves and Surfaces*

**Participants:** Kathlen Kohn, Bernd Sturmfels, Matthew Trager, Boris Bukh, Xavier Goaoc, Alfredo Hubard, Matthew Trager.

Visual events in computer vision are studied from the perspective of algebraic geometry. Given a sufficiently general curve or surface in 3-space, we consider the image or contour curve that arises by projecting from a viewpoint. Qualitative changes in that curve occur when the viewpoint crosses the visual event surface as illustrated in 3. We examine the components of this ruled surface, and observe that these coincide with the iterated singular loci of the coisotropic hypersurfaces associated with the original curve or surface. We derive formulas, due to Salmon and Petitjean, for the degrees of these surfaces, and show how to compute exact representations for all visual event surfaces using algebraic methods. This work has been published in [8].

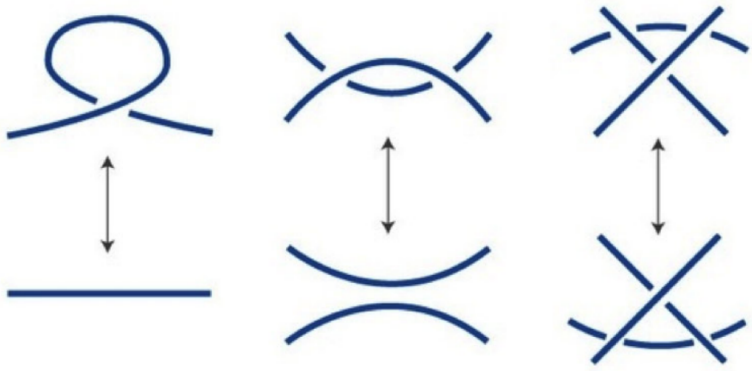

*Figure 3. Changing views of a curve correspond to Reidemeister moves. The viewpoint $z$ crosses the tangential surface (left), edge surface (middle), or trisecant surface (right).*

subsectionConsistent Sets of Lines with no Colorful Incidence

We consider incidences among colored sets of lines in $\mathbb{R}^d$ and examine whether the existence of certain concurrences between lines of $k$ colors force the existence of at least one concurrence between lines of $k+1$ colors. This question is relevant for problems in 3D reconstruction in computer vision such as the one illustrated in Figure 4. This work has been published in [12].

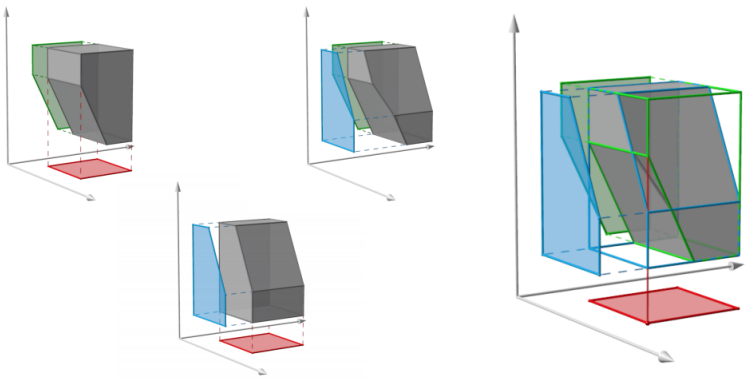

*Figure 4. Three silhouettes that are 2-consistent but not globally consistent for three orthogonal projections. Each of the first three figures shows a three-dimensional set that projects onto two of the three silhouettes. The fourth figure illustrates that no set can project simultaneously onto all three silhouettes: the highlighted red image point cannot be lifted in 3D, since no point that projects onto it belongs to the pre-images of both the blue and green silhouettes.*

### *7.1.4. On the Solvability of Viewing Graphs*

**Participants:** Matthew Trager, Brian Osserman, Jean Ponce.

A set of fundamental matrices relating pairs of cameras in some configuration can be represented as edges of a ” viewing graph ". Whether or not these fundamental matrices are generically sufficient to recover the global camera configuration depends on the structure of this graph. We study characterizations of ” solvable " viewing graphs, and present several new results that can be applied to determine which pairs of views may be used to recover all camera parameters. We also discuss strategies for verifying the solvability of a graph computationally. This work has been published in [21].

### *7.1.5. In Defense of Relative Multi-View Geometry*

**Participants:** Matthew Trager, Jean Ponce.

The idea of studying multi-view geometry and structure-from-motion problems *relative* to the scene and camera configurations, without appeal to external coordinate systems, dates back to the early days of modern geometric computer vision. Yet, it has a bad rap, the scene reconstructions obtained often being deemed as inaccurate despite careful implementations. The aim of this article is to correct this perception with a series of new results. In particular, we show that using a small subset of scene and image points to parameterize their relative configurations offers a natural coordinate-free formulation of Carlsson-Weinshall duality for arbitrary numbers of images. An example is shown in Figure 5. For three views, this approach also yields novel purely- and quasi-linear formulations of structure from motion using *reduced trilinearities*, without the complex polynomial constraints associated with trifocal tensors, revealing in passing the strong link between “3D” ($\mathbb{P}^3 \to \mathbb{P}^2$) and “2D” ($\mathbb{P}^2 \to \mathbb{P}^1$) models of trinocular vision. Finally, we demonstrate through preliminary experiments that the proposed relative reconstruction methods gives good results on real data. This works is available as a preprint [32].

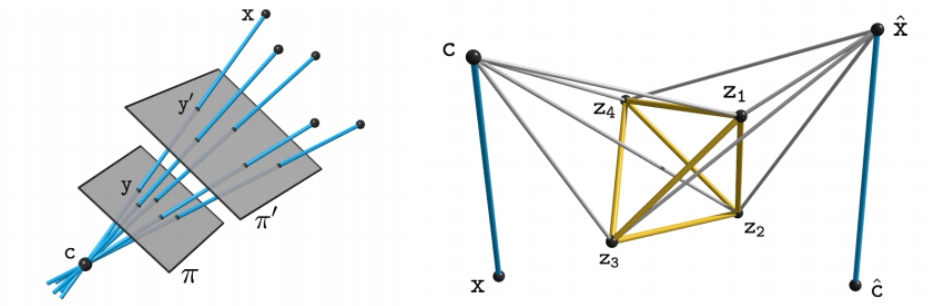


*Figure 5. Configurations.* ***Left****: Image point and viewing ray configurations are isomorphic and independent of the retinal plane.* ***Right****: Geometric Carlsson-Weinshall duality between scene point and pinhole configurations.*

### *7.1.6. Multigraded Cayley-Chow Forms*

**Participants:** Brian Osserman, Matthew Trager.

We introduce a theory of multigraded Cayley-Chow forms associated to subvarieties of products of projective spaces. Figure 6 illustrares some examples of projective speces. Two new phenomena arise: first, the construction turns out to require certain inequalities on the dimensions of projections; and second, in positive characteristic the multigraded Cayley-Chow forms can have higher multiplicities. The theory also provides a natural framework for understanding multifocal tensors in computer vision. This works is available as a preprint [30].

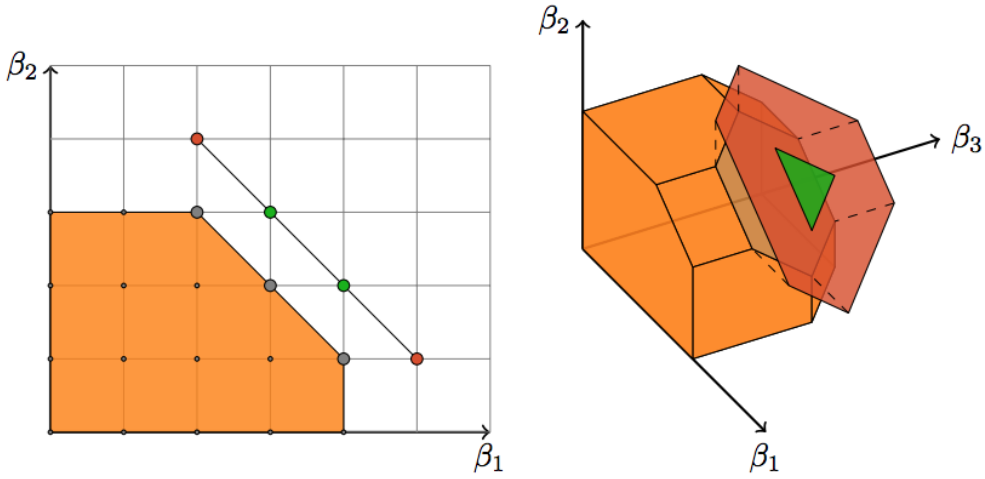


*Figure 6. Two polymatroids. The sets of bases (corresponding to our multidegree supports) are in gray; while the sets of circuits and of non-circuit 1-deficient vectors are in green and red, respectively.*

## 7.2. Category-level object and scene recognition

### *7.2.1. Detecting rare visual relations using analogies*

**Participants:** Julia Peyre, Cordelia Schmid, Ivan Laptev, Josef Sivic.

We seek to detect visual relations in images of the form of triplets t = (subject, predicate, object), such as "person riding dog", where training examples of the individual entities are available but their combinations are rare or unseen at training such as shown in Figure 7. This is an important set-up due to the combinatorial nature of visual relations : collecting sufficient training data for all possible triplets would be very hard. The contributions of this work are three-fold. First, we learn a representation of visual relations that combines (i) individual embeddings for subject, object and predicate together with (ii) a visual phrase embedding that represents the relation triplet. Second, we learn how to transfer visual phrase embeddings from existing training triplets to unseen test triplets using analogies between relations that involve similar objects. Third, we demonstrate the benefits of our approach on two challenging datasets involving rare and unseen relations : on HICO-DET, our model achieves significant improvement over a strong baseline, and we confirm this improvement on retrieval of unseen triplets on the UnRel rare relation dataset. This work, currently under review, can be found at [31].

### *7.2.2. Convolutional neural network architecture for geometric matching*

**Participants:** Ignacio Rocco, Relja Arandjelović, Josef Sivic.

In [9], we address the problem of determining correspondences between two images in agreement with a geometric model such as an affine, homography or thin-plate spline transformation, and estimating its parameters. The contributions of this work are threefold. First, we propose a convolutional neural network architecture for geometric matching. The architecture is based on three main components that mimic the standard steps of feature extraction, matching and simultaneous inlier detection and model parameter estimation, while being trainable end-to-end. Second, we demonstrate that the network parameters can be trained from synthetically generated imagery without the need for manual annotation and that our matching layer significantly increases generalization capabilities to never seen before images. Finally, we show that the same model can perform both instance-level and category-level matching giving state-of-the-art results on the challenging PF, TSS and Caltech-101 datasets.

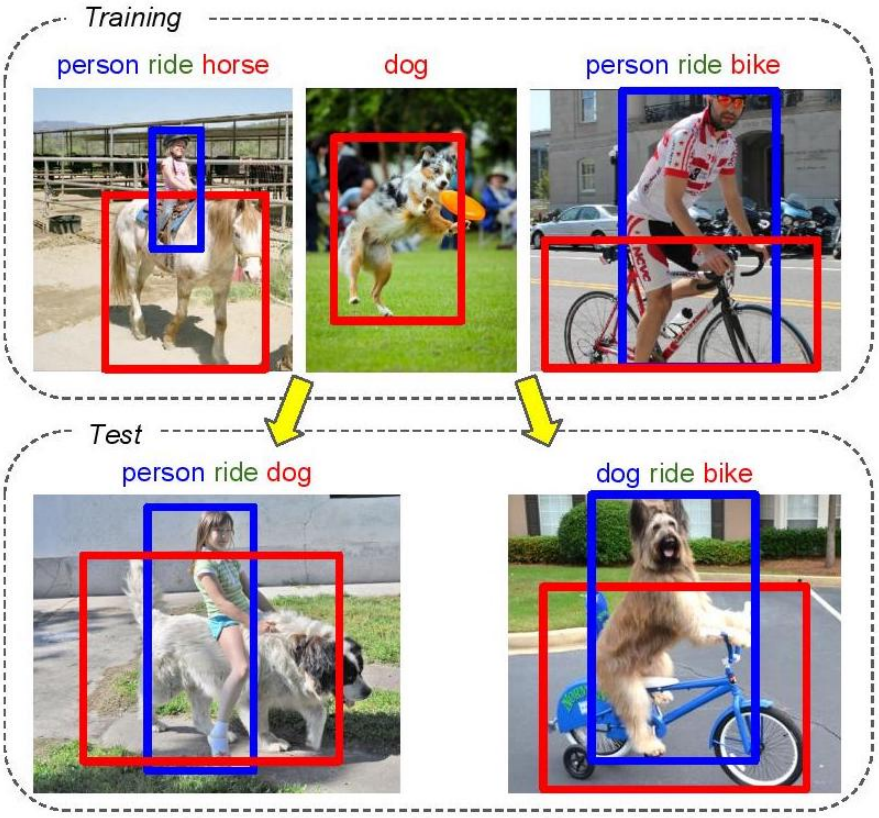


*Figure 7. Illustration of transfer by analogy from seen training triplets (e.g. “person ride horse") to unseen or rare ones (e.g. “person ride dog"))*

### 7.2.3. End-to-end weakly-supervised semantic alignment

**Participants:** Ignacio Rocco, Relja Arandjelović, Josef Sivic.

In [17], we tackle the task of semantic alignment where the goal is to compute dense semantic correspondence aligning two images depicting objects of the same category. This is a challenging task due to large intra-class variation, changes in viewpoint and background clutter. We present the following three principal contributions. First, we develop a convolutional neural network architecture for semantic alignment that is trainable in an end-to-end manner from weak image-level supervision in the form of matching image pairs. The outcome is that parameters are learnt from rich appearance variation present in different but semantically related images without the need for tedious manual annotation of correspondences at training time. Second, the main component of this architecture is a differentiable soft inlier scoring module, inspired by the RANSAC inlier scoring procedure, that computes the quality of the alignment based on only geometrically consistent correspondences thereby reducing the effect of background clutter. Third, we demonstrate that the proposed approach achieves state-of-the-art performance on multiple standard benchmarks for semantic alignment. Figure 8 presents some example results.

*Figure 8. Each row corresponds to one example and shows the (right) automatic semantic alignment of the (left) source and (middle) target images.*

### 7.2.4. Neighbourhood Consensus Networks

**Participants:** Ignacio Rocco, Mircea Cimpoi, Relja Arandjelović, Akihiko Torii, Tomas Pajdla, Josef Sivic.

In [18], we address the problem of finding reliable dense correspondences between a pair of images. This is a challenging task due to strong appearance differences between the corresponding scene elements and ambiguities generated by repetitive patterns. The contributions of this work are threefold. First, inspired by the classic idea of disambiguating feature matches using semi-local constraints, we develop an end-to-end trainable convolutional neural network architecture that identifies sets of spatially consistent matches by analyzing neighbourhood consensus patterns in the 4D space of all possible correspondences between a pair of images without the need for a global geometric model. Second, we demonstrate that the model can be trained effectively from weak supervision in the form of matching and non-matching image pairs without the need for costly manual annotation of point to point correspondences. Third, we show the proposed neighbourhood consensus network can be applied to a range of matching tasks including both category- and instance-level matching, obtaining the state-of-the-art results on the PF Pascal dataset and the InLoc indoor visual localization benchmark. Figure 9 shows the network architecture of the proposed Neighbourhood Consensus Network, that features 3 layers of 4D convolutions.

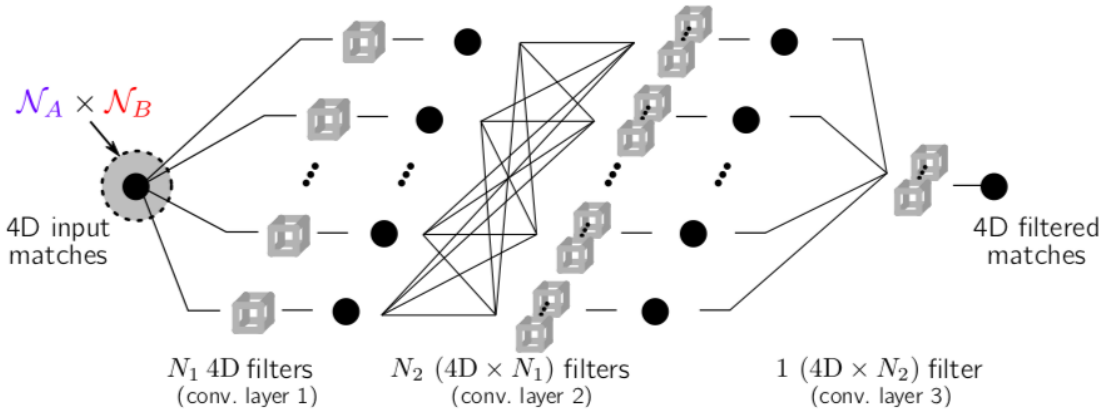


*Figure 9. A neighbourhood consensus CNN operates on the 4D space of feature matches. The first 4D convolutional layer filters span $\mathcal{N}_A \times \mathcal{N}_B$, the Cartesian product of local neighbourhoods $\mathcal{N}_A$ and $\mathcal{N}_B$ in images $A$ and $B$ respectively. The proposed 4D neighbourhood consensus CNN can learn to identify the matching patterns of reliable and unreliable matches, and filter the matches accordingly*

### *7.2.5. Compressing the Input for CNNs with the First-Order Scattering Transform*

**Participants:** Edouard Oyallon, Eugene Belilovsky, Sergey Zagoruyko, Michal Valko.

In [16], 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.

### *7.2.6. Exploring Weight Symmetry in Deep Neural Networks*

**Participants:** Xu Shell Hu, Sergey Zagoruyko, Nikos Komodakis.

In [27], we propose to impose symmetry in neural network parameters to improve parameter usage and make use of dedicated convolution and matrix multiplication routines. Due to significant reduction in the number of parameters as a result of the symmetry constraints, one would expect a dramatic drop in accuracy. Surprisingly, we show that this is not the case, and, depending on network size, symmetry can have little or no negative effect on network accuracy, especially in deep overparameterized networks. We propose several ways to impose local symmetry in recurrent and convolutional neural networks, and show that our symmetry parameterizations satisfy universal approximation property for single hidden layer networks. We extensively evaluate these parameterizations on CIFAR, ImageNet and language modeling datasets, showing significant benefits from the use of symmetry. For instance, our ResNet-101 with channel-wise symmetry has almost 25% less parameters and only 0.2% accuracy loss on ImageNet.

## 7.3. Image restoration, manipulation and enhancement

### *7.3.1. Neural Embedding of an Iterative Deconvolution Algorithm for Motion Blur Estimation and Removal*

**Participants:** Thomas Eboli, Jian Sun, Jean Ponce.

We introduce a new two-steps learning-based approach to motion blur estimation and removal decomposed into two trainable modules. A local linear motion model is estimated at each pixel using a first convolutional neural network (CNN) in a regression setting. It is then used to drive an algorithm that casts non-blind, non-uniform image deblurring as a least-squares problem regularized by natural image priors in the form of sparsity constraints. This problem is solved by combining the alternative direction method of multipliers with an iterative residual compensation algorithm, with a finite number of iterations embedded into a second CNN whose trainable parameters are deconvolution filters. The second network outputs the sharp image, and the two CNNs can be trained together in an end-to-end manner. Our experiments demonstrate that the proposed method is significantly faster than existing ones, and provides competitive results with the state of the art on synthetic and real data. This work is available as a pre-print[25] and an example is illustrated in Figure 10.

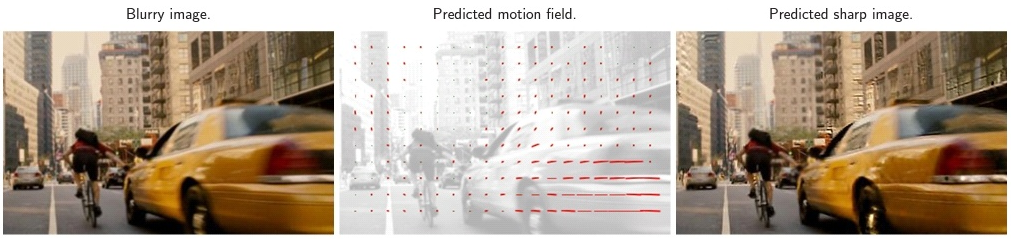


*Figure 10. From a blurry image, we first use CNN-based regressor to predict a motion field with local linear motions before using it in a trainable iterative residual compensation algorithm to restore the image.*

### *7.3.2. Deformable Kernel Networks for Joint Image Filtering*

**Participants:** Beomjun Kim, Jean Ponce, Bumsub Ham.

Joint image filters are used to transfer structural details from a guidance picture used as a prior to a target image, in tasks such as enhancing spatial resolution and suppressing noise. Previous methods based on convolutional neural networks (CNNs) combine nonlinear activations of spatially-invariant kernels to estimate structural details and regress the filtering result. In this paper, we instead learn explicitly sparse and spatially-variant kernels. We propose a CNN architecture and its efficient implementation, called the deformable kernel network (DKN), that outputs sets of neighbors and the corresponding weights adaptively for each pixel. The filtering

result is then computed as a weighted average. We also propose a fast version of DKN that runs about four times faster for an image of size 640 by 480. We demonstrate the effectiveness and flexibility of our models on the tasks of depth map upsampling, saliency map upsampling, cross-modality image restoration, texture removal, and semantic segmentation. In particular, we show that the weighted averaging process with sparsely sampled 3 by 3 kernels outperforms the state of the art by a significant margin. This works has been submitted to the IEEE Trans. on Pattern Analysis and Machine Intelligence and is available as a pre-print [28].

## 7.4. Human activity capture and classification

### *7.4.1. Learning a Text-Video Embedding from Incomplete and Heterogeneous Data*

**Participants:** Antoine Miech, Ivan Laptev, Josef Sivic.

Joint understanding of video and language is an active research area with many applications. Prior work in this domain typically relies on learning text-video embeddings. One difficulty with this approach, however, is the lack of large-scale annotated video-caption datasets for training. To address this issue, in [29] we aim at learning text-video embeddings from heterogeneous data sources. To this end, we propose a Mixture-of-Embedding-Experts (MEE) model with ability to handle missing input modalities during training. As a result, our framework can learn improved text-video embeddings simultaneously from image and video datasets. We also show the generalization of MEE to other input modalities such as face descriptors. We evaluate our method on the task of video retrieval and report results for the MPII Movie Description and MSR-VTT datasets. The proposed MEE model demonstrates significant improvements and outperforms previously reported methods on both text-to-video and video-to-text retrieval tasks. Figure 11 illustrates application of our method in text-to-video retrieval.

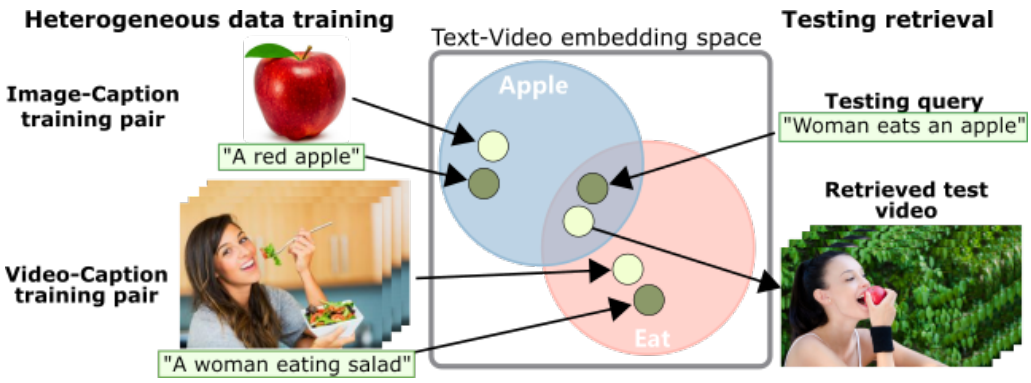


*Figure 11. We learn a text-video embedding from heterogenous (here Image-Text and Video-Text) data sources. At test time, we can query concepts learnt from both Image-Caption and Video-Caption training pair (e.g. the eating notion being learnt from video and the apple notion from image).*

### *7.4.2. A flexible model for training action localization with varying levels of supervision*

**Participants:** Guilhem Chéron, Jean-Baptiste Alayrac, Ivan Laptev, Cordelia Schmid.

Spatio-temporal action detection in videos is typically addressed in a fully-supervised setup with manual annotation of training videos required at every frame. Since such annotation is extremely tedious and prohibits scalability, there is a clear need to minimize the amount of manual supervision. In this work we propose a unifying framework that can handle and combine varying types of less-demanding weak supervision. Our model is based on discriminative clustering and integrates different types of supervision as constraints on the

optimization as illustrated in Figure 12. We investigate applications of such a model to training setups with alternative supervisory signals ranging from video-level class labels to the full per-frame annotation of action bounding boxes. Experiments on the challenging UCF101-24 and DALY datasets demonstrate competitive performance of our method at a fraction of supervision used by previous methods. The flexibility of our model enables joint learning from data with different levels of annotation. Experimental results demonstrate a significant gain by adding a few fully supervised examples to otherwise weakly labeled videos. This work has been published in [14].

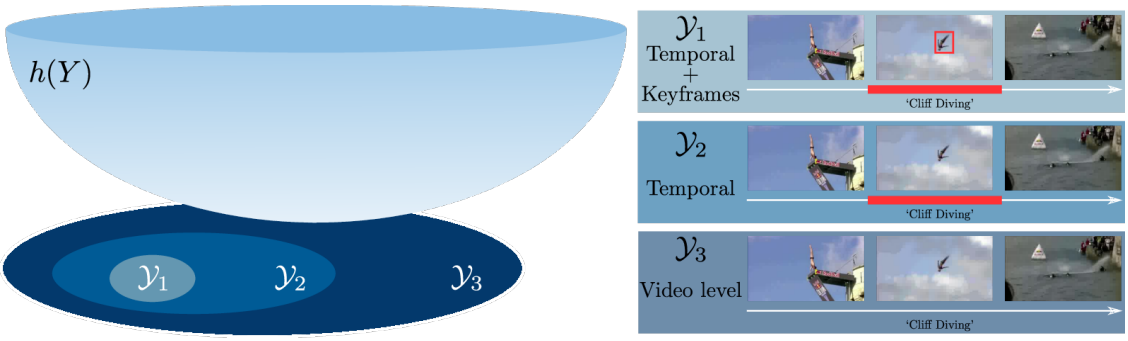


*Figure 12. Our method estimates a matrix $Y$ assigning human tracklets to action labels in training videos by optimizing an objective function $h(Y)$ under constraints $\mathcal{Y}_s$. Different types of supervision define particular constraints $\mathcal{Y}_s$ and do not affect the form of the objective function. The increasing level of supervision imposes stricter constraints, e.g. $\mathcal{Y}_1 \supset \mathcal{Y}_2 \supset \mathcal{Y}_3 \supset \mathcal{Y}_4$ as illustrated for the Cliff Diving example above.*

### *7.4.3. BodyNet: Volumetric Inference of 3D Human Body Shapes*

**Participants:** Gül Varol, Duygu Ceylan, Bryan Russell, Jimei Yang, Ersin Yumer, Ivan Laptev, Cordelia Schmid.

Human shape estimation is an important task for video editing, animation and fashion industry. Predicting 3D human body shape from natural images, however, is highly challenging due to factors such as variation in human bodies, clothing and viewpoint. Prior methods addressing this problem typically attempt to fit parametric body models with certain priors on pose and shape. In this work we argue for an alternative representation and propose BodyNet, a neural network for direct inference of volumetric body shape from a single image. BodyNet is an end-to-end trainable network that benefits from (i) a volumetric 3D loss, (ii) a multi-view re-projection loss, and (iii) intermediate supervision of 2D pose, 2D body part segmentation, and 3D pose. Each of them results in performance improvement as demonstrated by our experiments. To evaluate the method, we fit the SMPL model to our network output and show state-of-the-art results on the SURREAL and Unite the People datasets, outperforming recent approaches. Besides achieving state-of-the-art performance, our method also enables volumetric body-part segmentation. Figure 13 illustrates the volumetric outputs given two sample input images. This work has been published at ECCV 2018 [22].

### *7.4.4. Localizing Moments in Video with Temporal Language*

**Participants:** Lisa Anne Hendricks, Oliver Wang, Eli Schechtman, Josef Sivic, Trevor Darrell, Bryan Russell.

Localizing moments in a longer video via natural language queries is a new, challenging task at the intersection of language and video understanding. Though moment localization with natural language is similar to other language and vision tasks like natural language object retrieval in images, moment localization offers an interesting opportunity to model temporal dependencies and reasoning in text. In [15], we propose a new model that explicitly reasons about different temporal segments in a video, and shows that temporal context

*Figure 13. Our BodyNet predicts a volumetric 3D human body shape and 3D body parts from a single image. We show the input image, the predicted human voxels, and the predicted part voxels.*

is important for localizing phrases which include temporal language. To benchmark whether our model, and other recent video localization models, can effectively reason about temporal language, we collect the novel TEMPO-ral reasoning in video and language (TEMPO) dataset. Our dataset consists of two parts: a dataset with real videos and template sentences (TEMPO - Template Language) which allows for controlled studies on temporal language, and a human language dataset which consists of temporal sentences annotated by humans (TEMPO - Human Language).

### *7.4.5. The Pinocchio C++ library ? A fast and flexible implementation of rigid body dynamics algorithms and their analytical derivatives*

**Participants:** Justin Carpentier, Guilhem Saurel, Gabriele Buondonno, Joseph Mirabel, Florent Lamiraux, Olivier Stasse, Nicolas Mansard.

In this work, we introduce Pinocchio, an open-source software framework that implements rigid body dynamics algorithms and their analytical derivatives. Pinocchio does not only include standard algorithms employed in robotics (e.g., forward and inverse dynamics) but provides additional features essential for the control, the planning and the simulation of robots. In this paper, we describe these features and detail the programming patterns and design which make Pinocchio efficient. We evaluate the performances against RBDL, another framework with broad dissemination inside the robotics community. We also demonstrate how the source code generation embedded in Pinocchio outperforms other approaches of state of the art.

### *7.4.6. Modeling Spatio-Temporal Human Track Structure for Action Localization*

**Participants:** Guilhem Chéron, Anton Osokin, Ivan Laptev, Cordelia Schmid.

This paper [24] addresses spatio-temporal localization of human actions in video. In order to localize actions in time, we propose a recurrent localization network (RecLNet) designed to model the temporal structure of actions on the level of person tracks. Our model is trained to simultaneously recognize and localize action classes in time and is based on two layer gated recurrent units (GRU) applied separately to two streams, i.e. appearance and optical flow streams. When used together with state-of-the-art person detection and tracking, our model is shown to improve substantially spatio-temporal action localization in videos. The gain is shown to be mainly due to improved temporal localization as illustrated in Figure 14. We evaluate our method on two recent datasets for spatio-temporal action localization, UCF101-24 and DALY, demonstrating a significant improvement of the state of the art.

# 8. Bilateral Contracts and Grants with Industry

## 8.1. Bilateral Contracts with Industry

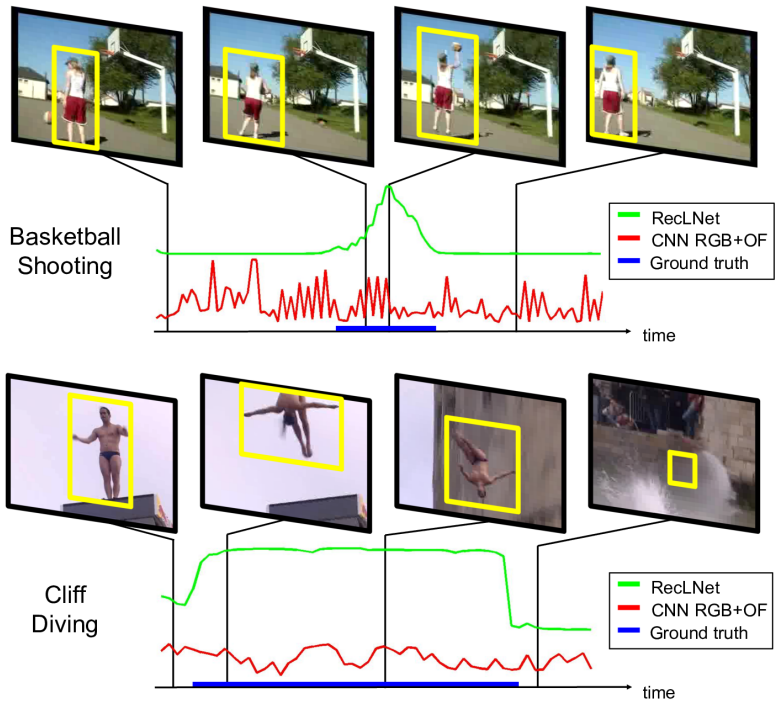


*Figure 14. Spatio-temporal action localization using a CNN baseline (red) and our RecLNet (green) both applied on the level of person tracks. Our approach provides accurate temporal boundaries when the action happens.*

### *8.1.1. MSR-Inria joint lab: Image and video mining for science and humanities (Inria)*

**Participants:** Guilhem Cheron, Ivan Laptev, Maxime Oquab, Jean Ponce, Josef Sivic, Cordelia Schmid [Inria Lear].

This collaborative project brings together the WILLOW and LEAR project-teams with MSR researchers in Cambridge and elsewhere. The concept builds on several ideas articulated in the 2020 Sciencea report, including the importance of data mining and machine learning in computational science. Rather than focusing only on natural sciences, however, we propose here to expand the breadth of e-science to include humanities and social sciences. The project we propose will focus on fundamental computer science research in computer vision and machine learning, and its application to archaeology, cultural heritage preservation, environmental science, and sociology, and it will be validated by collaborations with researchers and practitioners in these fields.

In October 2013 a new agreement has been signed for 2013-2017 with the research focus on automatic understanding of dynamic video content. Recent studies predict that by 2018 video will account for 80-90% of traffic on the Internet. Automatic understanding and interpretation of video content is a key enabling factor for a range of practical applications such as organizing and searching home videos or content aware video advertising. For example, interpreting videos of "making a birthday cake" or "planting a tree" could provide effective means for advertising products in local grocery stores or garden centers. The goal of this project is to perform fundamental computer science research in computer vision and machine learning in order to enhance the current capabilities to automatically understand, search and organize dynamic video content.

In 2018 a new agreement has been signed with a new focus on video understanding for personal assistants. The scientific objectives are to develop models, representations and learning algorithms for (i) automatic understanding of task-driven complex human activities from videos narrated with natural language in order to (ii) give people instructions in a new environment via an augmented reality device such as the Microsoft HoloLens. Besides the clear scientific interest of automatically understanding human activities in video streams, the main high-impact motivation of this project it to develop virtual assistants that may guide a child through simple games to improve his/her manipulation and language skills; help an elderly person to achieve everyday tasks; or facilitate the training of a new worker for highly-specialized machinery maintenance.

### *8.1.2. Louis Vuitton/ENS chair on artificial intelligence*

**Participants:** Ivan Laptev, Jean Ponce, Josef Sivic.

The scientific chair Louis Vuitton - École normale supérieure in Artificial Intelligence has been created in 2017 and inaugurated on April 12, 2018 by the ENS Director Marc Mézard and the LV CEO Michael Burke. The goal of the chair is to establish a close collaboration between LV and ENS in the area of Artificial Intelligence. The chair enjoys the generous annual contribution of 200K Euros provided by LV in support of research activities in statistical learning and computer vision. In particular, the chair supports the costs of researchers, students, missions, computational resources as well as seminars and meetings, including the two days of meeting annually organized by LV and ENS. During 2018 ENS and LV have organized several joint meetings with the participation of researchers from SIERRA and WILLOW teams. The chair has also supported the hiring of one PhD student at the WILLOW team, missions to conferences and international research labs as well as data collection for research projects.

## 8.2. Bilateral Grants with Industry

### *8.2.1. Facebook AI Research Paris: Weakly-supervised interpretation of image and video data (Inria)*

**Participants:** Jean Ponce, Minsu Cho, Ivan Laptev, Josef Sivic.

We will develop in this project (Facebook gift) new models of image and video content, as well as new recognition architectures and algorithms, to address the problem of understanding the visual content of images and videos using weak forms of supervision, such as the fact that multiple images contain instances of the same objects, or the textual information available in television or film scripts.

### *8.2.2. Google: Learning to annotate videos from movie scripts (Inria)*

**Participants:** Josef Sivic, Ivan Laptev, Jean Ponce.

The goal of this project is to automatically generate annotations of complex dynamic events in video. We wish to deal with events involving multiple people interacting with each other, objects and the scene, for example people at a party in a house. The goal is to generate structured annotations going beyond simple text tags. Examples include entire text sentences describing the video content as well as bounding boxes or segmentations spatially and temporally localizing the described objects and people in video. This is an extremely challenging task due to large intra-class variation of human actions. We propose to learn joint video and text representations enabling such annotation capabilities from feature length movies with coarsely aligned shooting scripts. Building on our previous work in this area, we aim to develop structured representations of video and associated text enabling to reason both spatially and temporally about scenes, objects and people as well as their interactions. Automatic understanding and interpretation of video content is a key-enabling factor for a range of practical applications such as content-aware advertising or search. Novel video and text representations are needed to enable breakthrough in this area.

### *8.2.3. Google: Structured learning from video and natural language (Inria)*

**Participants:** Simon Lacoste-Julien, Ivan Laptev, Josef Sivic.

People can easily learn how to change a flat tire of a car or assemble an IKEA shelve by observing other people doing the same task, for example, by watching a narrated instruction video. In addition, they can easily perform the same task in a different context, for example, at their home. This involves advanced visual intelligence abilities such as recognition of objects and their function as well as interpreting sequences of human actions that achieve a specific task. However, currently there is no artificial system with a similar cognitive visual competence. The goal of this proposal is to develop models, representations and learning algorithms for automatic understanding of complex human activities from videos narrated with natural language.

# 9. Partnerships and Cooperations

## 9.1. National Initiatives

### *9.1.1. DGA - RAPID project DRAAF*

**Participant:** Ivan Laptev.

DGA DRAAF is a two-year collaborative effort with University of Caen (F. Jurie) and the industrial partner EVITECH (P. Bernas) focused on modelling and recognition of violent behaviour in surveillance videos. The project aims to develop image recognition models and algorithms to automatically detect weapons, gestures and actions using recent advances in computer vision and deep learning to provide an affordable real-time solution reducing effects of threats in public places.

## 9.2. European Initiatives

### *9.2.1. European Research Council (ERC) Starting Grant: “Activia" - Ivan Laptev*

**Participant:** Ivan Laptev.

WILLOW will be funded in part from 2013 to 2018 by the ERC Starting Grant "Activia" awarded to Ivan Laptev by the European Research Council.

‘Computer vision is concerned with the automated interpretation of images and video streams. Today’s research is (mostly) aimed at answering queries such as ’Is this a picture of a dog?’, (classification) or sometimes ’Find the dog in this photo’ (detection). While categorisation and detection are useful for many tasks, inferring correct class labels is not the final answer to visual recognition. The categories and locations of objects do not provide direct understanding of their function i.e., how things work, what they can be used for, or how they can act and react. Such an understanding, however, would be highly desirable to answer currently unsolvable queries such as ’Am I in danger?’ or ’What can happen in this scene?’. Solving such queries is the aim of this proposal. My goal is to uncover the functional properties of objects and the purpose of actions by addressing visual recognition from a different and yet unexplored perspective. The main novelty of this proposal is to leverage observations of people, i.e., their actions and interactions to automatically learn the use, the purpose and the function of objects and scenes from visual data. The project is timely as it builds upon the two key recent technological advances: (a) the immense progress in visual recognition of objects, scenes and human actions achieved in the last ten years, as well as (b) the emergence of a massive amount of public image and video data now available to train visual models. ACTIVIA addresses fundamental research issues in automated interpretation of dynamic visual scenes, but its results are expected to serve as a basis for ground-breaking technological advances in practical applications. The recognition of functional properties and intentions as explored in this project will directly support high-impact applications such as detection of abnormal events, which are likely to revolutionise today’s approaches to crime protection, hazard prevention, elderly care, and many others.’

### *9.2.2. European Research Council (ERC) Starting Grant: "Leap" - Josef Sivic*

**Participant:** Josef Sivic.

The contract has begun on Nov 1st 2014. WILLOW will be funded in part from 2014 to 2018 by the ERC Starting Grant "Leap" awarded to Josef Sivic by the European Research Council.

‘People constantly draw on past visual experiences to anticipate future events and better understand, navigate, and interact with their environment, for example, when seeing an angry dog or a quickly approaching car. Currently there is no artificial system with a similar level of visual analysis and prediction capabilities. LEAP is a first step in that direction, leveraging the emerging collective visual memory formed by the unprecedented amount of visual data available in public archives, on the Internet and from surveillance or personal cameras - a complex evolving net of dynamic scenes, distributed across many different data sources, and equipped with plentiful but noisy and incomplete metadata. The goal of this project is to analyze dynamic patterns in this shared visual experience in order (i) to find and quantify their trends; and (ii) learn to predict future events in dynamic scenes. With ever expanding computational resources and this extraordinary data, the main scientific challenge is now to invent new and powerful models adapted to its scale and its spatio-temporal, distributed and dynamic nature. To address this challenge, we will first design new models that generalize across different data sources, where scenes are captured under vastly different imaging conditions such as camera viewpoint, temporal sampling, illumination or resolution. Next, we will develop a framework for finding, describing and quantifying trends that involve measuring long-term changes in many related scenes. Finally, we will develop a methodology and tools for synthesizing complex future predictions from aligned past visual experiences. Our models will be automatically learnt from large-scale, distributed, and asynchronous visual data, coming from different sources and with different forms of readily-available but noisy and incomplete metadata such as text, speech, geotags, scene depth (stereo sensors), or gaze and body motion (wearable sensors). Breakthrough progress on these problems would have profound implications on our everyday lives as well as science and commerce, with safer cars that anticipate the behavior of pedestrians on streets; tools that help doctors monitor, diagnose and predict patients’ health; and smart glasses that help people react in unfamiliar situations enabled by the advances from this project.’

## 9.3. International Initiatives

### *9.3.1. IMPACT: Intelligent machine perception*

**Participants:** Josef Sivic, Jean Ponce, Ivan Laptev.

IMPACT is a 5-year collaborative project with Czech Technical University, Center for Robotics, Informatics and Cybernetics (CIIRC) (2017-2022). The IMPACT project focuses on fundamental and applied research in computer vision, machine learning and robotics to develop machines that learn to perceive, reason, navigate and interact with complex dynamic environments. For example, people easily learn how to change a flat tire of a car or perform resuscitation by observing other people doing the same task. This involves advanced visual intelligence abilities such as interpreting sequences of human actions that manipulate objects to achieve a specific task. Currently, however, there is no artificial system with a similar level of cognitive visual competence. Breakthrough progress in intelligent machine perception will have profound implications on our everyday lives as well as science and commerce, with smart assistive robots that automatically learn new skills from the Internet, safer cars that autonomously navigate in difficult changing conditions, or intelligent glasses that help people navigate never seen before environments.

### *9.3.2. Associate team GAYA*

**Participants:** Jean Ponce, Matthew Trager.

GAYA is a joint research team bringing together two Inria project-teams (Thoth, Grenoble and WILLOW, Paris) and Carnegie Mellon University, USA. It focuses on two research themes: (i) semantic structured interpretation of videos, and (ii) studying the geometric properties of object shapes to enhance state-of-the-art object recognition approaches.

Interpreting videos semantically in a general setting, involving various types of video content like home video clips, news broadcasts, feature films, which contain a lot of clutter, non-rigid motion, many "actors" performing actions, person-object and person-person interactions, varying viewpoints, is challenging. This task is being examined increasingly over the past decade, with the availability of large video resources, e.g., YouTube. Despite this progress, an effective video representation for recognizing actions is still missing. To address this critical challenge, we propose a joint optimization framework, wherein we learn the video representation and also develop models for action recognition. Specifically, we aim to exploit the spatio-temporal relations among pixels in a video through graphical models and novel deep learning feature representations.

The second research theme explores geometric aspects of computer vision, in particular how to model three-dimensional objects from their two-dimensional projections, and how the appearance of these objects evolves with changes in viewpoint. Beyond its theoretical interest, this work is critical for developing object recognition algorithms that take into account the three-dimensional nature of the visual world and go beyond the template-matching approaches dominant today. Duality is an important concept in this area, and we are investigating its application to the construction of visual hulls as well as the characterization of the topology of image contours using the Gauss map. Existing results are essentially limited to the Euclidean setting, and we are investigating their generalization to the general projective case.

Partners: CMU (Deva Ramanan, Martial Hebert, Abhinav Gupta, Gunnar Sigurdsson), Inria Thoth (Cordelia Schmid, Karteek Alahari, Pavel Tokmakov).

## 9.4. International Research Visitors

### *9.4.1. Visits of International Scientists*

Alexei Efros (Professor, UC Berkeley, USA) visited Willow during May-June. Ramazan Cinbis (Middle East Technical University) and David Fouhey (University of Michigan) visited Willow in July-August and September-November, respectively. Akihiko Torii (Tokyo Institute of Technology) spent sabbatical at Willow from Apr to August 2018. Finally, Pierre-Yves Masse (post-doc, Czech Technical University) spent 50% of his time at Sierra (F. Bach) and Willow teams as a visiting post-doc within the framework of collaboration with the Intelligent Machine Perception project lead by J. Sivic at the Czech Technical University in Prague.

# 10. Dissemination

## 10.1. Promoting Scientific Activities

### *10.1.1. Scientific Events Organisation*

*10.1.1.1. General Chair, Scientific Chair*

- I. Laptev was program co-chair of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

*10.1.1.2. Member of the Organizing Committees*

- G. Varol and Y. Hasson are co-organizers of “Women in Computer Vision Workshop” at European Conference on Computer Vision (ECCV), 2018.
- I. Laptev and J. Sivic were co-organizers of “Fine-grained instructional video understanding workshop" at IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

### *10.1.2. Scientific Events Selection*

*10.1.2.1. Area chairs*

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018 (J. Sivic).
- European Conference on Computer Vision (ECCV), 2018 (I. Laptev, J. Sivic).

*10.1.2.2. Member of the Conference Program Committees / reviewer*

- IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018 (J.-B. Alayrac, M. Oquab, R. Rezende, I. Rocco, G. Varol).
- European Conference on Computer Vision (ECCV), 2018 (A. Miech, G. Varol, I. Rocco, S. Zagaryuko).
- Neural Information Processing Systems (NIPS), 2018 (J. Sivic).
- Asian Conference on Computer Vision (ACCV), 2018 (G. Varol).

### *10.1.3. Journal*

*10.1.3.1. Member of the Editorial Boards*

- International Journal of Computer Vision (I. Laptev, J. Ponce, J. Sivic).
- IEEE Transactions on Pattern Analysis and Machine Intelligence (I. Laptev, J. Sivic).
- Foundations and Trends in Computer Graphics and Vision (J. Ponce).

*10.1.3.2. Reviewer - Reviewing Activities*

- International Journal of Computer Vision (G. Cheron, M. Trager, G. Varol).
- IEEE Transactions on Pattern Analysis and Machine Intelligence (J.-B. Alayrac, G. Cheron, M. Trager, G. Varol).
- IEEE Transactions on Circuits and Systems for Video Technology (G. Varol).

### *10.1.4. Other*

- J. Sivic is senior fellow of the Neural Computation and Adaptive Perception program of the Canadian Institute of Advanced Research.

### *10.1.5. Invited Talks*

- I. Laptev, Keynote, ICCVG, Warsaw, September, 2018.
- I. Laptev, Invited talk, EPIC Workshop in conjunction with ECCV’18, Munich, September 2018.
- I. Laptev, Tutorial, BMVC’18, Newcastle, September 2018.

- I. Laptev, Invited talk, Workshop on Brave New Ideas for Video Understanding, in conjunction with CVPR'18, Salt Lake City, June 2018.
- I. Laptev, Invited talk, Journee AI, SAFRAN, Paris, June 2018.
- I. Laptev, Invited talk, Integrating Vision and Language, Tartu, March 2018.
- I. Laptev, Keynote, 36th Annual Swedish Symposium on Image Analysis, Stockholm, March 2018.
- A. Miech, Invited talk, LSCP-ENS seminar, Paris, March 2018.
- A. Miech, Invited talk, Google, Paris, July 2018.
- A. Miech, Invited talk, Google, Mountain View, July 2018.
- A. Miech, Invited talk, Paris ML Meetup, University of Bristol Computer Vision Seminar, Bristol, UK, November 2018.
- J. Ponce, Artificial Intelligence, French-American Joint Committee Meeting on Science and Technology, College de France, March 9, 2018.
- J. Ponce, Weakly supervised structure discovery in images and videos, Intelligent robots: autonomy and vision, NYU Abu Dhabi, March 13, 2018.
- J. Ponce, From vision and robotics to artificial intelligence, Robotics AI: Data science vs motion intelligence symposium co-organized by the French and German Academies of Sciences, Sep. 5, 2018.
- J. Ponce, Shape, contours, cameras and eyes, Workshop in honor of Jan Koenderink, UC Berkeley, UC berkeley, Oct. 24.
- J. Ponce, Weakly supervised structure discovery in images and videos, NYU Tandon School of Engineering, New York, Nov. 2, 2018.
- J. Ponce, Computer vision and visual recognition: historical perspective, new results and challenges, University of Zagreb, Zagreb, Croatia, Nov. 9, 2018.
- J. Sivic, Invited talk, Paris Sciences et Data, PSL, 02/2018.
- J. Sivic, Invited talk, AIME@CZ - Czech workshop on applied mathematics in engineering, Czech Technical University, 02/2018.
- J. Sivic, Invited talk, Deep Learning Workshop, CVPR 2018, Salt Lake City, June 2018.
- J. Sivic, Invited talk, Landmark Recognition Workshop, CVPR 2018, Salt Lake City, June 2018.
- J. Sivic, Seminar, UC Berkeley, June 2018.
- J. Sivic, Invited talk, Workshop on YouTube-8M Large-Scale Video Understanding, ECCV 2018, Munich, September 2018.
- J. Sivic, Invited talk, Prague Informatics Seminar, Charles University, Prague, April 2018.
- G. Varol, Invited talk, MPI for Informatics, Saarbrucken, Germany 11/2018.
- G. Varol, Invited talk, Istanbul Technical University, Istanbul, Turkey 10/2018.
- G. Varol, Invited talk, Deep Learning Meetup at Station F, Paris, France 09/2018.
- G. Varol, Invited talk, BNP Paribas - Prairie Summer School, Paris, France 08/2018.
- G. Varol, Invited talk, CTU Center for Machine Perception, Prague, Czech Republic 06/2018.
- G. Varol, Invited talk, MPI for Intelligent Systems, Tubingen, Germany 04/2018.

### *10.1.6. Leadership within the Scientific Community*

- Member of the advisory board for the IBM Watson AI Xprize (J. Ponce).
- Member of the steering committee of France AI (J. Ponce).
- Member, advisory board, Computer Vision Foundation (J. Sivic).

### *10.1.7. Scientific Expertise*

- J. Ponce, coordinator of the AI theme for the joint French-American Committee on Science and Technology, 2018-.

### *10.1.8. Research Administration*

- Member, Bureau du comité des projets, Inria, Paris (J. Ponce)
- Member, Scientific academic council, PSL Research University (J. Ponce)
- Member, Research representative committee, PSL Research University (J. Ponce).
- Member of Inria Commission de developpement technologique (CDT), 2012-2018 (J. Sivic).
- Member of the Hiring Committe for the tenure track position at CentraleSupelec (I. Laptev).
- Member of the Hiring Committee for Professor of Computer Vision at CentraleSupelec (I. Laptev).

## 10.2. Teaching - Supervision - Juries

### *10.2.1. Teaching*

- Master : M. Aubry, K. Alahari, I. Laptev and J. Sivic "Introduction to computer vision", M1, Ecole normale superieure, 36h.
- Master : I. Laptev, J. Ponce and J. Sivic (together with C. Schmid, Inria Grenoble), "Object recognition and computer vision", M2, Ecole normale superieure, and MVA, Ecole normale superieure de Cachan, 36h.
- J. Ponce co-organized the PRAIRIE AI Summer School, Grenoble, 2018, which brought together 200 participants representing 44 different nationalities, and selected from 700 applications, with 60% students, 15% academics , and 25% industrials. 25% of these participants were women.

### *10.2.2. Supervision*

PhD : Jean-Baptiste Alayrac, “Structured Learning from Videos and Language”, graduated in September 2018, I. Laptev, J. Sivic and S. Lacoste-Julien (Inria SIERRA / U. Montreal).

PhD : Guilhem Cheron, “Structured modeling and recognition of human actions in video”, graduated in Dec 2018, I. Laptev and C. Schmid.

PhD : Maxime Oquab, “Convolutional neural networks: towards less supervision for visual recognition”, defended on 26 January 2018, L. Bottou (Facebook AI Research), I. Laptev and J. Sivic.

PhD : Matthew Trager, “Cameras, Shapes, and Contours: Geometric Models in Computer Vision”, graduated in 2018, J. Ponce and M. Hebert (CMU).

PhD : Tuang Hung VU, “Learning visual models for person detection and action prediction”, graduated in 2018 , I. Laptev.

PhD in progress : Vo Van Huy, started in Dec 2018, J. Ponce.

PhD in progress : Robin Strudel, “Learning and transferring complex robot skills from human demonstrations”, started in Oct 2018, I. Laptev, C. Schmid and J. Sivic.

PhD in progress : Yann Labbe, “Generalizing robotic sensorimotor skills to new tasks and environments", started in Oct 2018, J. Sivic and I. Laptev.

PhD in progress : Thomas Eboli, started in Oct 2017, J. Ponce.

PhD in progress : Zongmian Li, “Learning to manipulate objects from instructional videos”, started in Oct 2017, I. Laptev, J. Sivic and N. Mansard (LAAS/CNRS, Toulouse).

PhD in progress : Yana Hasson, started in Nov 2017, I. Laptev and C. Schmid.

PhD in progress : Dmitry Zhukov, “Learning from instruction videos for personal assistants”, started in Oct 2017, I. Laptev and J. Sivic.

PhD in progress : Ignacio Rocco, “Estimating correspondence between images via convolutional neural networks”, started in Jan 2017, J. Sivic, R. Arandjelovic (Google DeepMind).

PhD in progress : Antoine Miech, “Understanding long-term temporal structure of videos”, started in Oct 2016, I. Laptev, J. Sivic, P. Bojanowski (Facebook AI Research).

PhD in progress : Gul Varol, “Deep learning methods for video interpretation”, started in Oct 2015, I. Laptev, C. Schmid.

PhD in progress : Julia Peyre, “Learning to reason about scenes from images and language”, started in Oct 2015, C. Schmid, I. Laptev, J. Sivic.

PhD in progress : Theophile Dalens, “Learning to analyze and reconstruct architectural scenes”, starting in Jan 2015, M. Aubry and J. Sivic.

### *10.2.3. Juries*

PhD thesis committee:

- Taylor MORDAN, Sorbonne Universite, France, 2018, (J. Sivic, rapporteur).
- Stephane LATHUILIERE, Universite Grenoble Alpes, France, 2018, (J. Sivic, rapporteur).
- Tuan-Hung VU, PSL University, France, 2018, (J. Sivic, examinateur).
- Siddhartha Chandra, CentraleSupelec, France, 2018, (I. Laptev examinateur).
- Sergey Zagoruyko, l?Universite Paris-Est, France, 2018, (I. Laptev examinateur).
- Joris Guerin, Arts et Metiers ParisTech, France, 2018, (I. Laptev examinateur).
- Guilhem Cheron, PSL, France, 2018, (J. Ponce examinateur).
- Pavel Tokmakov, Universite Grenoble Alpes, France, 2018, (J. Ponce examinateur).

## 10.3. Popularization

### *10.3.1. Articles and contents*

- J. Ponce was the subject of an article in the photography magazine Polka. He was also interviewed by France Culture, Le Monde, Paris-Match, Québec Science, and Science et Vie.
- J. Ponce was interviewed by the Académie des Sciences working group on Artificial Intelligence on Oct. 2, 2018.

### *10.3.2. Interventions*

- J. Ponce participated in round tables about AI at the France Culture Forum in Paris, March 1, 2018, at the BNP Parisbas summer school Aug. 23, 2018, at the French embassy in Berlin on Nov. 6, 2018, at the Erasme-Descartes conference in Paris on Nov. 15, 2018, and at the AI summit in New York City on Dec. 6, 2018.
- J. Ponce gave general audience lectures at the X-IA meeting in Palaiseau, Oct. 8, 2018 and at the Institut français de Croatie in Zagreb, Croatia, Nov. 9, 2018.

### *10.3.3. Creation of media or tools for science outreach*

- I. Laptev, Y. Hasson, S. Allayen, T. Eboli, I. Kalyevath, Y. Labbe, R. Strudel, G. Varol, D. Zhukov, Presentation of computer vision for high-school students, Inria, December 2018.

# 11. Bibliography

## Publications of the year

### Doctoral Dissertations and Habilitation Theses

[1] J.-B. ALAYRAC.*Structured Learning from Videos and Language*, Ecole normale supérieure - ENS PARIS, September 2018, https://hal.inria.fr/tel-01885412

[2] G. CHÉRON.*Structured modeling and recognition of human actions in video*, Ecole normale supérieure - ENS PARIS, December 2018, https://hal.inria.fr/tel-01975247

[3] M. OQUAB.*Convolutional neural networks: towards less supervision for visual recognition*, Ecole Normale Supérieure (ENS) ; ED 386 : École doctorale de sciences mathématiques de Paris centre, UPMC, January 2018, https://hal.inria.fr/tel-01803967

[4] M. TRAGER.*Cameras, Shapes, and Contours: Geometric Models in Computer Vision*, Ecole Normale Superieure de Paris - ENS Paris, July 2018, https://hal.inria.fr/tel-01856415

[5] T.-H. VU.*Learning visual models for person detection and action prediction*, Ecole Normale Superieure de Paris - ENS Paris, September 2018, https://hal.inria.fr/tel-01861455

### Articles in International Peer-Reviewed Journal

[6] B. HAM, M. CHO, J. PONCE.*Robust Guided Image Filtering Using Nonconvex Potentials*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", January 2018, vol. Vol. 40, n$^{o}$ No. 1, p. 291-207, Accepted pending minor revision [*DOI :* 10.1109/TPAMI.2017.2669034], https://hal.archives-ouvertes.fr/hal-01279857

[7] B. HAM, M. CHO, C. SCHMID, J. PONCE.*Proposal Flow: Semantic Correspondences from Object Proposals*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", July 2018, vol. 40, n$^{o}$ 7, p. 1711-1725 [*DOI :* 10.1109/TPAMI.2017.2724510], https://hal.inria.fr/hal-01644132

[8] K. KOHN, B. STURMFELS, M. TRAGER.*Changing Views on Curves and Surfaces*, in "Acta Mathematica Vietnamica", 2018, https://arxiv.org/abs/1707.01877 - 31 pages [*DOI :* 10.1007/s40306-017-0240-1], https://hal.inria.fr/hal-01676208

[9] I. ROCCO, R. ARANDJELOVIĆ, J. SIVIC.*Convolutional neural network architecture for geometric matching*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", 2018, p. 1-14 [*DOI :* 10.1109/TPAMI.2018.2865351], https://hal.archives-ouvertes.fr/hal-01859616

[10] G. VAROL, I. LAPTEV, C. SCHMID.*Long-term Temporal Convolutions for Action Recognition*, in "IEEE Transactions on Pattern Analysis and Machine Intelligence", June 2018, vol. 40, n$^{o}$ 6, p. 1510-1517, https://arxiv.org/abs/1604.04494 [*DOI :* 10.1109/TPAMI.2017.2712608], https://hal.inria.fr/hal-01241518

[11] Y. ZHANG, Y. SU, J. YANG, J. PONCE, H. KONG.*When Dijkstra meets vanishing point: a stereo vision approach for road detection*, in "IEEE Transactions on Image Processing", 2018, p. 1-12, https://hal.archives-ouvertes.fr/hal-01678548

### International Conferences with Proceedings

[12] B. BUKH, X. GOAOC, A. HUBARD, M. TRAGER.*Consistent Sets of Lines with no Colorful Incidence*, in "SoCG 2018 - 34thInternational Symposium on Computational Geometry", Budapest, Hungary, E. SPECKMANN, C. D. TÓTH (editors), Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, June 2018, p. 1-20, https://arxiv.org/abs/1803.06267 , https://hal-upec-upem.archives-ouvertes.fr/hal-01744125

[13] J. CARPENTIER, G. SAUREL, G. BUONDONNO, J. MIRABEL, F. LAMIRAUX, O. STASSE, N. MANSARD.*The Pinocchio C++ library – A fast and flexible implementation of rigid body dynamics*

*algorithms and their analytical derivatives*, in "SII 2019 - International Symposium on System Integrations", Paris, France, January 2019, https://hal.laas.fr/hal-01866228

[14] G. CHÉRON, J.-B. ALAYRAC, I. LAPTEV, C. SCHMID.*A flexible model for training action localization with varying levels of supervision*, in "NIPS 2018 - 32nd Conference on Neural Information Processing Systems", Montréal, Canada, December 2018, p. 1-17, https://arxiv.org/abs/1806.11328 , https://hal.inria.fr/hal-01937002

[15] L. A. HENDRICKS, O. WANG, E. SHECHTMAN, J. SIVIC, T. DARRELL, B. RUSSELL.*Localizing Moments in Video with Temporal Language*, in "Empirical Methods in Natural Language Processing (EMNLP)", Brussels, Belgium, October 2018, https://arxiv.org/abs/1809.01337 - EMNLP 2018, https://hal.archives-ouvertes.fr/hal-01976945

[16] 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

[17] I. ROCCO, R. ARANDJELOVIĆ, J. SIVIC.*End-to-end weakly-supervised semantic alignment*, in "CVPR 2018 - IEEE Conference on Computer Vision and Pattern Recognition", Salt Lake City, UT, United States, June 2018, p. 1-9, https://hal.archives-ouvertes.fr/hal-01859628

[18] I. ROCCO, M. CIMPOI, R. ARANDJELOVIĆ, A. TORII, T. PAJDLA, J. SIVIC.*Neighbourhood Consensus Networks*, in "32nd Conference on Neural Information Processing Systems (NIPS 2018)", Montréal, Canada, December 2018, https://arxiv.org/abs/1810.10510 , https://hal.archives-ouvertes.fr/hal-01905474

[19] T. SATTLER, W. MADDERN, C. TOFT, A. TORII, L. HAMMARSTRAND, E. STENBORG, D. SAFARI, M. OKUTOMI, M. POLLEFEYS, J. SIVIC, F. KAHL, T. PAJDLA.*Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions*, in "IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2018)", Salt Lake City, UT, United States, June 2018, https://arxiv.org/abs/1707.09092 , https://hal.archives-ouvertes.fr/hal-01859660

[20] H. TAIRA, M. OKUTOMI, T. SATTLER, M. CIMPOI, M. POLLEFEYS, J. SIVIC, T. PAJDLA, A. TORII.*InLoc: Indoor Visual Localization with Dense Matching and View Synthesis*, in "CVPR 2018 - IEEE Conference on Computer Vision and Pattern Recognition", Salt Lake City, United States, June 2018, https://arxiv.org/abs/1803.10368 , https://hal.archives-ouvertes.fr/hal-01859637

[21] M. TRAGER, B. OSSERMAN, J. PONCE.*On the Solvability of Viewing Graphs*, in "European Conference on Computer Vision 2018 (ECCV 2018)", Munich, Germany, September 2018, https://hal.inria.fr/hal-01856159

[22] G. VAROL, D. CEYLAN, B. RUSSELL, J. YANG, E. YUMER, I. LAPTEV, C. SCHMID.*BodyNet: Volumetric Inference of 3D Human Body Shapes*, in "ECCV 2018 - 15th European Conference on Computer Vision", Munich, Germany, September 2018, p. 1-27, https://arxiv.org/abs/1804.04875 , https://hal.inria.fr/hal-01852169

## Other Publications

[23] R. BUDHIRAJA, J. CARPENTIER, N. MANSARD.*Dynamics Consensus between Centroidal and Whole-Body Models for Locomotion of Legged Robots*, September 2018, Accepted for IEEE International Conference on Robotics and Automation 2019, https://hal.laas.fr/hal-01875031

[24] G. CHÉRON, A. OSOKIN, I. LAPTEV, C. SCHMID.*Modeling Spatio-Temporal Human Track Structure for Action Localization*, January 2019, https://arxiv.org/abs/1806.11008 - working paper or preprint, https://hal.inria.fr/hal-01979583

[25] T. EBOLI, J. SUN, J. PONCE.*Neural Embedding of an Iterative Deconvolution Algorithm for Motion Blur Estimation and Removal*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01857177

[26] M. HAHN, N. RUIZ, J.-B. ALAYRAC, I. LAPTEV, J. M. REHG.*Learning to Localize and Align Fine-Grained Actions to Sparse Instructions*, January 2019, https://arxiv.org/abs/1809.08381 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01979719

[27] X. S. HU, S. ZAGORUYKO, N. KOMODAKIS.*Exploring Weight Symmetry in Deep Neural Networks*, December 2018, https://arxiv.org/abs/1812.11027 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01978633

[28] B. KIM, J. PONCE, B. HAM.*Deformable Kernel Networks for Joint Image Filtering*, October 2018, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01857016

[29] A. MIECH, I. LAPTEV, J. SIVIC.*Learning a Text-Video Embedding from Incomplete and Heterogeneous Data*, January 2019, working paper or preprint, https://hal.archives-ouvertes.fr/hal-01975102

[30] B. OSSERMAN, M. TRAGER.*Multigraded Cayley-Chow Forms*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01856190

[31] J. PEYRE, I. LAPTEV, C. SCHMID, J. SIVIC.*Detecting rare visual relations using analogies*, January 2019, https://arxiv.org/abs/1812.05736 - working paper or preprint, https://hal.archives-ouvertes.fr/hal-01975760

[32] M. TRAGER, J. PONCE.*In Defense of Relative Multi-View Geometry*, August 2018, working paper or preprint, https://hal.inria.fr/hal-01676732

[33] T.-H. VU, A. OSOKIN, I. LAPTEV.*Tube-CNN: Modeling temporal evolution of appearance for object detection in video*, January 2019, https://arxiv.org/abs/1812.02619 - 13 pages, 8 figures, technical report, https://hal.archives-ouvertes.fr/hal-01980339